International Textbook of Diabetes Mellitus
Section Editors

Knut Borch-Johnsen MD, DMSc
Professor of Quality in Health Care
Medical Director, Holbæk Hospital
Holbaek, Denmark

Mark Cooper MD, PhD, FRACP
Deputy Director and Chief Scientific Officer
Baker IDI Heart and Diabetes Institute
Melbourne, VIC, Australia

Stefano Del Prato MD
Professor of Endocrinology and Metabolism
School of Medicine, University of Pisa;
Chief, Section of Diabetes and Metabolic Diseases
University of Pisa
Pisa, Italy

Robert H. Eckel MD
Charles A. Boettcher Endowed Chair in Atherosclerosis
Professor of Medicine
Division of Endocrinology, Metabolism and Diabetes, and Cardiology
Professor of Physiology and Biophysics
University of Colorado School of Medicine
Aurora, CO, USA

W. Timothy Garvey MD, FACE
Butterworth Professor and Chair,
Department of Nutrition Sciences
University of Alabama at Birmingham
Director, UAB Diabetes Research Center
GRECC Investigator and Staff Physician
Birmingham VA Medical Center
Birmingham, AL, USA

Rury R. Holman FRCP
Professor of Diabetic Medicine
Director, University of Oxford Diabetes Trials Unit
Oxford Centre for Diabetes, Endocrinology and Metabolism
Churchill Hospital
University of Oxford
Oxford, UK

Steven E. Kahn MB, ChB
Professor of Medicine
Division of Metabolism, Endocrinology and Nutrition
Director, Diabetes Research Center
University of Washington
VA Puget Sound Health Care System
Seattle, WA, USA

James B. Meigs MD
Massachusetts General Hospital
Boston, MA, USA

Eric Ravussin PhD
Associate Executive Director, Obesity & Diabetes
Pennington Biomedical Research Center
Baton Rouge, LA, USA

Jay S. Skyler MD, MACP
Professor of Medicine, Pediatrics, & Psychology
Division of Endocrinology, Diabetes, & Metabolism
Deputy Director—Diabetes Research Institute
University of Miami Miller School of Medicine
Miami, FL, USA

Hannele Yki-Järvinen MD, FRCP
Professor of Medicine
University of Helsinki;
Minerva Foundation Institute for Medical Research
Helsinki, Finland

Bernard Zinman CM, MD, FRCPC, FACP
Director, Leadership Sinai Centre for Diabetes
Sam and Judy Pencer Family Chair in Diabetes Research
Senior Scientist, Samuel Lunenfeld Research Institute
Professor of Medicine, University of Toronto Mount Sinai Hospital
Toronto, ON, Canada
Contents

Contributors, xi
Preface, xxiii

Section I: Epidemiology
1 Classification of diabetes mellitus and other categories of glucose intolerance, 3
   Dianna J. Magliano, Paul Zimmet and Jonathan E. Shaw
2 Epidemiology and risk factors for type 1 diabetes mellitus, 17
   Andrea K. Steck and Marian J. Rewers
3 Epidemiology and geography of type 2 diabetes mellitus, 29
   Qing Qiao, Desmond E. Williams, Giuseppina Imperatore, K.M. Venkat Narayan and Jaakko Tuomilehto

Section II: Physiology of pancreatic function
4 Development and maintenance of the islet β cell, 55
   Teresa L. Mastracci, Raghavendra G. Mirmira and Ryan M. Anderson
5 Pancreatic morphology in normal and diabetic states, 69
   Rebecca L. Hull and Steven E. Kahn
6 Insulin gene expression and biosynthesis, 82
   Vincent Poitout, Roland Stein and Christopher J. Rhodes
7 β-Cell biology of insulin secretion, 96
   Susumu Seino, Tadao Shibusaki and Kohtaro Minami
8 Normal β-cell function, 108
   Andrea Natali, Stefano Del Prato and Andrea Mari
9 Neuropeptides and islet hormone secretion, 125
   Bo Ahrén
10 Biosynthesis, secretion, and action of glucagon, 136
   Pierre J. Lefèvre
11 Incretin physiology in health and disease, 145
   Juris J. Meier and Michael A. Nauck

Section III: Normal physiology of insulin action
12 Mechanisms of insulin signal transduction, 163
   Paul R. Langlais, Lawrence J. Mandarino and W. Timothy Garvey
13 Regulation of glucose metabolism in liver, 193
   Shawn C. Burgess
14 Insulin actions in vivo: glucose metabolism, 211
   Ele Ferrannini and Ralph A. DeFronzo
15 Measuring insulin action in vivo, 234
   Richard N. Bergman and Marilyn Ader
16 Protein metabolism in health and diabetes, 250
   Elena Volpi and Jared M. Dickinson
17 Lipid and lipoprotein metabolism, hypolipidemic agents, and therapeutic goals, 262
   Gisette Reyes-Soffer and Henry N. Ginsberg
18 Metabolomics: applications in type 2 diabetes mellitus and insulin resistance, 275
   Sean H. Adams and John W. Newman

Section IV: NAFLD, NASH and non-traditional associations with diabetes
19 Pathogenesis of nonalcoholic fatty liver disease (NAFLD), 283
   Hannele Yki-Järvinen
20 Treatment of nonalcoholic fatty liver disease (NAFLD) and nonalcoholic steatohepatitis (NASH), 292
   Fernando Bril, Kwame Ntim, Romina Lomonaco and Kenneth Cusi
21 Type 2 diabetes and cancer, 306
   Emily Jane Gallagher, Ruslan Novosyadlyy and Derek LeRoith
22 Diabetes and sleep apnea, 316
   Abd A. Tahrani
**Section V: Pathogenesis**

23 The insulin resistance syndrome, 339  
   Naveed Sattar, Paul Welsh and David Preiss

24 β-Cell mass and function in human type 2 diabetes, 354  
   Piero Marchetti and Ele Ferrannini

25 Pathogenesis of type 2 diabetes mellitus, 371  
   Ralph A. DeFronzo

26 The genetics of type 2 diabetes, 401  
   Emma Ahlqvist and Leif Groop

27 Glucose toxicity, 413  
   Hannele Yki-Järvinen and Donald A. McClain

28 Monogenic disorders of the β cell, 426  
   Ali J. Chakera and Andrew T. Hattersley

29 Immunopathogenesis of type 1 diabetes in Western society, 442  
   Alberto Pugliese

30 Molecular genetics of type 1 diabetes, 454  
   M. Ann Kelly, Simon D. Rees, Anthony H. Barnett and Stephen C. Bain

**Section VI: Obesity**

31 The relationship between obesity and type 2 diabetes—the role of gut factors, 469  
   Daniel S. Hsia and William T. Cefalu

32 The role of energy metabolism in the regulation of energy balance, 479  
   Charmaine S. Tam and Eric Ravussin

33 Treatment of obesity: lifestyle and pharmacotherapy, 489  
   Arne Astrup

34 Treatment of obesity: bariatric surgery, 505  
   Karl J. Neff, Ele Ferrannini and Carol W. le Roux

35 Animal models of obesity and type 2 diabetes, 519  
   David A. York, Mie Jung Park-York and Timothy Gilbertson

36 The role of the hypothalamus in the maintenance of energy balance and peripheral glucose control, 529  
   Luis Varela and Tamas L. Horvath

**Section VII: Clinical trial evidence in diabetes**

37 Prevention of type 1 diabetes, 541  
   Jay S. Skyler

38 Prevention of type 2 diabetes, 550  
   Thomas Yates and Melanie J. Davies

39 Prevention of diabetic microvascular complications, 564  
   Alice P.S. Kong and Juliana C.N. Chan

**Section VIII: Management of diabetes: diet, exercise and drugs**

40 Dietary management of diabetes mellitus in Europe and North America, 577  
   Jim I. Mann and Alexandra Chisholm

41 Exercise, 589  
   Henrik Galbo and Erik A. Richter

42 Sulfonylureas and meglitinides: insights into physiology and translational clinical utility, 615  
   Harold E. Lebovitz and Arne Melander

43 Metformin and other biguanides: pharmacology and therapeutic usage, 641  
   Susan L. Samson and Alan J. Garber

44 PPAR agonists in the treatment of diabetes, 657  
   Sunder Mudaliar and Robert R. Henry

45 α-Glucosidase inhibitors, 673  
   Josée Leroux-Stewart, Rémi Rabasa-Lhoret and Jean-Louis Chiasson

46 Combination therapy in type 2 diabetes mellitus, 686  
   Muhammad Abdul-Ghani, Ralph A. DeFronzo and Roy Eldor

47 New drugs for the treatment of diabetes mellitus, 709  
   Clifford J. Bailey

48 Incretin-based therapies, 726  
   Michael A. Nauck and Juris J. Meier

49 Insulin pumps, 745  
   Dace L. Trence and Irl B. Hirsch

50 Innovative therapies in diabetes: colesevelam and bromocriptine, 758  
   Shamsa Ali and Vivian Fonseca

**Section IX: Management of diabetes: monitoring and other interventions**

51 Implantable pumps and artificial and bio-artificial pancreas system, 767  
   William V. Tamborlane and Miladys M. Palau-Collazo

52 Pancreas and islet transplantation, 774  
   R. Paul Robertson

53 Hypoglycemia and other complications of insulin therapy, 783  
   Stephanie A. Amiel
54 Diabetic ketoacidosis and hyperosmolar state, 799
Abbas E. Kitabchi, Guillermo E. Umpierrez and Mary Beth Murphy

Section X: Management of diabetes: glucose and lipid monitoring, specific subgroups

55 Type 2 diabetes in obese adolescents: pathophysiology and treatment, 817
Grace Kim and Sonia Caprio

56 Diabetes in pregnancy, 823
Yoel Toledano, Eran Hadar and Moshe Hod

57 Aging and diabetes mellitus, 836
Hermes Florez, Jorge Ruiz and Medha Munshi

58 Psychological problems and management of patients with diabetes mellitus, 846
Alan M. Delamater, Ashley Marchante and Jennifer Hernandez

59 Glycated hemoglobin, serum proteins, and other markers as tools for monitoring, 853
Susan M. Gerber, Douglas F. Stickle, Intekhab Ahmed and Serge A. Jabbour

Section XI: Complications of diabetes: general and microvascular

60 Pathogenesis of diabetic microvascular complications, 875
Josephine M. Forbes and Brooke Harcourt

61 Diabetic retinopathy and other ocular complications, 889
Gavin S. Tan and Tien Yin Wong

62 Diabetic nephropathy, 911

63 Pathology of human diabetic neuropathy, 926
M. Mojaddidi, A.K. Sharma and Rayaz A. Malik

64 Autonomic neuropathy, 939
Chinmay S. Marathe, Liza Phillips and Michael Horowitz

65 Connective tissue disorders in diabetes, 953
Arlan L. Rosenbloom

66 The diabetic foot, 964
Frank L. Bowling and Andrew J.M. Boulton

67 Erectile dysfunction in diabetes mellitus, 975
Michael H. Cummings and Rajesh Rajendran

68 Periodontal disease and diabetes mellitus, 988
Wenche S. Borgnakke and Robert J. Genco

Section XII: Complications of diabetes: macrovascular

69 Epidemiology of macrovascular disease and hypertension in diabetes mellitus, 1007
Jaakko Tuomilehto, Daiva Rastenye, Qing Qiao, Noël C. Barengo and Karl Matz

70 Atherogenesis, coronary heart disease and insulin resistance syndrome in diabetes, 1031
Barak Zafir and Jorge Plutzky

71 Endothelial function and metabolic syndrome, 1046
Xinpu Chen and Sridevi Devaraj

72 Hemostatic abnormalities in diabetes mellitus, 1051
R.A. Ajjan and Peter J. Grant

73 Clinical features and treatment of coronary heart disease in diabetes, 1064
Eberhard Standl, Michael Erbach, Markolf Hanefeld, Oliver Schnell, Bernd Stratmann and Diethelm Tschoepe

74 Arterial hypertension in diabetes: etiology and treatment, 1079
Colleen Flynn and George Bakris

75 Peripheral vascular and cerebrovascular disease in diabetes mellitus, 1091
Debabrata Mukherjee

Section XIII: Diabetes and public health

76 The diabetes challenge: from human and social rights to the empowerment of people with diabetes, 1105
C.T. Di Iorio, Fabrizio Carinci and M. Massi Benedetti

77 The economics of diabetes care: a global perspective, 1113
Rhys Williams

Index, i
Contributors

Muhammad Abdul-Ghani MD, PhD  
Associate Professor  
Diabetes Division  
Department of Medicine  
The University of Texas Health Science Center at San Antonio (UTHSCSA)  
San Antonio, TX, USA

Sean H. Adams MS, PhD  
Arkansas Children’s Nutrition Center, and Section of Developmental Nutrition  
Department of Pediatrics  
University of Arkansas for Medical Sciences  
Little Rock, AR, USA

Marilyn Ader PhD  
Associate Director, Cedars-Sinai Diabetes and Obesity Research Institute  
Associate Professor, Department of Biomedical Sciences  
Cedars-Sinai  
Los Angeles, CA, USA

Emma Ahlqvist PhD  
Assistant Professor  
Department of Clinical Sciences Malmö  
Lund University  
University Hospital Skåne  
Malmö, Sweden

Intekhab Ahmed MD, FACP, FACE  
Associate Professor of Medicine  
Program Director Endocrine Fellowship  
Division of Endocrinology, Diabetes and Metabolic Diseases  
Jefferson Medical College, Thomas Jefferson University  
Philadelphia, PA, USA

Bo Ahrén MD, PhD  
Professor  
Department of Clinical Sciences  
Lund University  
Lund, Sweden

R.A. Ajjan FRCP, MMedSci, PhD  
Associate Professor/Consultant in Diabetes and Endocrinology  
Division of Cardiovascular and Diabetes Research  
University of Leeds/Leeds Teaching Hospitals Trust  
Leeds, UK

Shamsa Ali MD  
Assistant Professor of Endocrinology  
Tulane University Health Sciences Center  
New Orleans, LA, USA

Stephanie A. Amiel MD, FRCP  
RD Lawrence Professor of Diabetic Medicine  
Joint Head of Division of Diabetes & Nutritional Sciences  
King’s College London  
London, UK

Ryan M. Anderson PhD  
Assistant Professor  
Department of Pediatrics  
Indiana University School of Medicine,  
Indianapolis, IN, USA

Arne Astrup MD, DMSc  
Director and Professor  
Department of Nutrition, Exercise and Sports  
Faculty of Science  
University of Copenhagen  
Copenhagen, Denmark

Clifford J. Bailey PhD, FRCP (Edin), FRCPath  
Professor of Clinical Science  
School of Life and Health Sciences  
Aston University  
Birmingham, UK

Stephen C. Bain MA, MD, FRCP  
Professor of Medicine (Diabetes)  
Institute of Life Science, College of Medicine  
Swansea University  
Swansea, UK

Scott T. Baker MB.BS, PhD  
Endocrinologist  
Endocrine Centre & Department of Medicine  
Austin Health;  
University of Melbourne  
Melbourne, VIC, Australia
George Bakris MD, FASH, FASN
Professor of Medicine
Director, ASH Comprehensive Hypertension Center
The University of Chicago
Chicago, IL, USA

Noël C. Barengo
Adjunct Professor
Department of Public Health
University of Helsinki
Helsinki, Finland

Anthony H. Barnett BSc (Hons), MD, FRCP
Consultant Physician and Emeritus Professor of Medicine
Clinical and Experimental Medicine
College of Medicine and Dentistry
University of Birmingham;
Diabetes Centre
Heart of England NHS Foundation Trust
Birmingham, UK

Richard N. Bergman PhD
Alfred Jay Firestein Chair in Diabetes Research
Director, Cedars-Sinai Diabetes and Obesity Research Institute
Professor, Department of Biomedical Sciences
Cedars-Sinai
Los Angeles, CA, USA

Wenche S. Borgnakke DDS, MPH, PhD
Senior Research Associate in Health Sciences
Department of Periodontics and Oral Medicine
University of Michigan School of Dentistry
Ann Arbor, MI, USA

Andrew J.M. Boulton MD, DSc (Hon), FRCP
President, European Association for the Study of Diabetes;
Professor of Medicine, University of Manchester;
Consultant Physician, Manchester Royal Infirmary
Manchester, UK;
Visiting Professor, University of Miami
Miami, FL, USA

Frank L. Bowling BSc(Hons), MSc, PhD, FFPM, RCPS (Glasg)
Clinical Research Fellow
University of Manchester
Manchester Royal Infirmary
Department of Diabetes & Vascular Surgery
Manchester, UK

Fernando Bril MD
Postdoctoral Associate
Division of Endocrinology, Diabetes & Metabolism
University of Florida
Gainesville, FL, USA

Shawn C. Burgess PhD
Associate Professor
Division Head, Metabolic Mechanisms of Disease
Advanced Imaging Research Center and Department of Pharmacology
University of Texas Southwestern Medical Center
Dallas, TX, USA

Sonia Caprio MD
Department of Pediatrics
Yale University School of Medicine
New Haven, CT, USA

Fabrizio Carinci MS
Senior Biostatistician
Serectrix snc
Pescara, Italy

William T. Cefalu MD
Executive Director
Douglas L. Manship, Sr. Professor of Diabetes
Pennington Biomedical Research Center
Louisiana State University
Baton Rouge, LA, USA

Ali J. Chakera MBChB, BMedSci (Hons)
Research Fellow
University of Exeter Medical School;
Royal Devon and Exeter Hospital
Exeter, UK

Juliana C.N. Chan MD
Professor of Medicine and Therapeutics
Department of Medicine and Therapeutics
Li Ka Shing Institute of Health Sciences
The Chinese University of Hong Kong
Hong Kong SAR, China

Xinpu Chen PhD
Postdoctoral Research Associate Baylor College of Medicine
Texas Children’s Hospital and Baylor College of Medicine
Houston, TX, USA

Jean-Louis Chiasson MD
Professor of Medicine
Division of Endocrinology
Centre hospitalier de l’Université de Montréal (CHUM);
Research Centre, Centre Hospitalier de l’Université de Montréal, Montréal (CRCHUM),
Department of Medicine
Université de Montréal
Montréal, Canada

Alexandra Chisholm PhD
Senior Research Fellow
Department of Human Nutrition
University of Otago
Dunedin, New Zealand
Contributors

Michael H. Cummings MD, FRCP
Professor of Diabetes and Endocrinology
Academic Department of Diabetes and Endocrinology
Queen Alexandra Hospitals Portsmouth NHS Trust
Portsmouth, UK

Kenneth Cusi MD, FACP, FACE
Professor of Medicine
Chief, Division of Endocrinology, Diabetes and Metabolism Division
University of Florida
Gainesville, FL, USA

Melanie J. Davies MD, FRCP, FRCGP
Professor of Diabetes Medicine
Leicester Diabetes Centre
College of Medicine, Biological Sciences and Psychology
University of Leicester;
National Institute for Health Research Leicester-Loughborough Diet
Lifestyle and Physical Activity
Biomedical Research Unit
Leicester, UK

Ralph A. DeFronzo MD
Professor of Medicine and Division Chief, Diabetes
Deputy Director, Texas Diabetes Institute
The University of Texas Health Science Center at San Antonio (UTHSCSA)
San Antonio, TX, USA

Alan M. Delamater PhD
Professor of Pediatrics and Psychology
Department of Pediatrics
University of Miami Miller School of Medicine
Miami, FL, USA

Stefano Del Prato MD
Professor of Endocrinology and Metabolism
School of Medicine, University of Pisa;
Chief, Section of Diabetes and Metabolic Diseases
University of Pisa
Pisa, Italy

Sridevi Devaraj PhD, DABCC, FACB
Medical Director of Clinical Chemistry and POCT;
Texas Children’s Hospital and Health Centers;
Professor of Pathology and Immunology
Baylor College of Medicine;
Director, Clinical Chemistry Fellowship
Associate Director, Translation, Texas Children’s Microbiome Center
Suite, Houston, TX, USA

Jared M. Dickinson PhD
Assistant Professor
School of Nutrition and Health Promotion
Healthy Lifestyles Research Center
Exercise Science and Health Promotion
Arizona State University
Phoenix, AZ, USA

Elif I. Ekinci MBBS, FRACP, PhD
Director of Diabetes
Endocrine Centre & Department of Medicine
Austin Health;
University of Melbourne
Melbourne, VIC, Australia

Roy Eldor MD, PhD
Prin. Scientist
LSD Clinical Research Diabetes and Endocrinology
Merck Research Laboratories
Rahway, NJ, USA

Michael Erbach MD
Manager Medical Education
Siarc Institute
Baierrbrunn, Germany

Ele Ferrannini MD
Professor of Internal Medicine
Department of Clinical and Experimental Medicine
University of Pisa
Pisa, Italy

Hermes Florez MD, PhD, MPH
Professor of Medicine & Public Health Sciences
Chief, Division of Gerontology & Geriatric Medicine
Director, Division of Epidemiology & Population Health
University of Miami Miller School of Medicine
Director, GRECC-Miami VA Healthcare System
Miami, FL, USA

Colleen Flynn, MD
Assistant Professor of Medicine
Department of Medicine, Section of Endocrinology, Diabetes and Metabolism
The University of Chicago
Chicago, IL, USA

Vivian Fonseca MD
Professor of Medicine & Pharmacology
Chief, Section of Endocrinology
Tulane University Health Sciences Center
New Orleans, LA, USA

Josephine M. Forbes BSc, PhD
Principal Research Fellow and Professor of Medicine
Group Leader, Glycation and Diabetes
Program Chief, Cardiovascular and Metabolic Diseases
Mater Research Institute, The University of Queensland
Brisbane, QLD Australia

Henrik Galbo MD
Professor
Institute of Inflammation Research
Department of Rheumatology, Rigshospitalet
University of Copenhagen
Copenhagen, Denmark
Emily Jane Gallagher, MD  
Assistant Professor  
Division of Endocrinology, Diabetes and Bone Diseases  
Icahn School of Medicine at Mount Sinai (ISMMS)  
New York, NY, USA

Alan J. Garber MD, PhD, FACE  
Professor  
Departments of Medicine,  
Molecular and Cellular Biology  
Baylor College of Medicine  
Houston, TX, USA

W. Timothy Garvey MD, FACE  
Butterworth Professor and Chair,  
Department of Nutrition Sciences  
University of Alabama at Birmingham  
Director, UAB Diabetes Research Center  
GRECC Investigator and Staff Physician  
Birmingham VA Medical Center  
Birmingham, AL, USA

Robert J. Genco DDS, PhD  
Distinguished Professor of Oral Biology and Microbiology  
State University of New York at Buffalo  
Buffalo, NY, USA

Susan M. Gerber MD  
Assistant Professor of Medicine  
Cooper Medical School of Rowan University  
Cooper University Health Care  
Camden, NJ, USA

Timothy Gilbertson PhD  
Professor  
Department of Biology  
Utah State University  
Logan, UT, USA

Henry N. Ginsberg MD  
Irving Professor of Medicine  
Department of Medicine  
Director, Irving Institute for Clinical and Translational Research  
Columbia University College of Physicians and Surgeons  
New York, NY, USA

Peter J. Grant MD, FRCPI, FMedSci  
Professor of Medicine  
Honorary Consultant Physician, ITHT  
Division of Cardiovascular and Diabetes Research  
Multidisciplinary Cardiovascular Research Centre  
University of Leeds  
Leeds, UK

Leif Groop MD  
Professor of Endocrinology  
Department of Clinical Sciences Malmö  
Lund University  
University Hospital Skåne  
Malmö, Sweden;  
Finnish Institute for Molecular Medicine (FIMM)  
University of Helsinki  
Helsinki, Finland

Eran Hadar MD  
Division of Maternal Fetal Medicine  
Helen Schneider Hospital for Women  
Rabin Medical Center  
Petah-Tiqva, Israel

Markolf Hanefeld MD, PhD  
Professor of Internal Medicine  
Study Centre Prof. Hanefeld  
Section Endocrinology and Metabolism  
GWT-Technical University  
Dresden, Germany

Brooke Harcourt BSc (Hons), PhD  
Mater Medical Research Institute  
South Brisbane, Australia;  
Department of Medicine and Immunology  
Monash University  
Melbourne, VIC, Australia

Andrew T. Hattersley DM, FRCP, FRS  
Professor of Molecular Medicine  
University of Exeter Medical School;  
Royal Devon and Exeter Hospital  
Exeter, UK

Robert R. Henry MD  
Chief, Section of Diabetes/Metabolism  
VA San Diego HealthCare System;  
Professor of Medicine  
University of California at San Diego,  
San Diego, CA, USA

Jennifer Hernandez MS  
Research Associate  
Department of Pediatrics  
University of Miami Miller School of Medicine  
Miami, FL, USA

Irl B. Hirsch MD  
Professor of Medicine  
University of Washington School of Medicine  
University of Washington Medical Center-Roosevelt  
Seattle, WA, USA
Moshe Hod MD  
Professor of Obstetrics and Gynecology  
Director, Division of Maternal Fetal Medicine  
Helen Schneider Hospital for Women  
Rabin Medical Center  
Petah-Tiqva, Israel

Michael Horowitz MBBS, PhD, FRACP  
Director, Endocrine & Metabolic Unit  
Royal Adelaide Hospital;  
Professor of Medicine and Leader  
Centre of Research Excellence (CRE) in Translating Nutritional Science to Good Health  
Discipline of Medicine  
University of Adelaide  
Adelaide, SA, Australia

Tamas L. Horvath DVM, PhD  
Professor & Chair  
Section of Comparative Medicine  
Jean and David W. Wallace Professor of Biomedical Research  
Professor of Neurobiology and Ob/Gyn  
Director, Yale Program in Integrative Cell Signaling and Neurobiology of Metabolism  
New Haven, CT, USA

Daniel S. Hsia MD  
Assistant Professor  
Pennington Biomedical Research Center  
Louisiana State University  
Baton Rouge, LA, USA

Rebecca L. Hull PhD  
Research Associate Professor  
Division of Metabolism, Endocrinology and Nutrition  
Department of Medicine,  
VA Puget Sound Health Care System and University of Washington  
Seattle, WA, USA

Giuseppina Imperatore  
Centers for Disease Control and Prevention  
National Center for Chronic Disease Prevention and Health Promotion  
Atlanta, GA, USA

Concetta Tania Di Iorio MPH  
Legal Expert  
Serecittix snc  
Pescara, Italy

Serge A. Jabbour MD, FACP, FACE  
Professor of Medicine  
Director, Division of Endocrinology, Diabetes & Metabolic Diseases  
Jefferson Medical College  
Thomas Jefferson University  
Philadelphia, PA, USA

George Jerums MBBS MD FRACP  
Endocrinologist  
Endocrine Centre & Department of Medicine  
Austin Health;  
University of Melbourne  
Melbourne, VIC, Australia

Steven E. Kahn MB, ChB  
Professor of Medicine  
Division of Metabolism, Endocrinology and Nutrition  
Director, Diabetes Research Center  
University of Washington  
VA Puget Sound Health Care System  
Seattle, WA, USA

M. Ann Kelly BSc (Hons), PhD  
Senior Scientist  
Clinical and Experimental Medicine, College of Medicine and Dentistry  
University of Birmingham;  
Diabetes Centre  
Heart of England NHS Foundation Trust  
Birmingham, UK

Grace Kim MD  
Post-doctoral Fellow  
Department of Pediatrics  
Yale University School of Medicine  
New Haven, CT, USA

Abbas E. Kitabchi PhD, MD, FACP, FACE  
Professor of Medicine and Molecular Sciences  
Maston K. Callison Professor of Medicine  
The University of Tennessee Health Science Center  
Memphis, TN, USA

Alice P.S. Kong MBChB, MD, MRCP, FRCP  
Associate Professor  
Department of Medicine and Therapeutics  
Li Ka Shing Institute of Health Sciences  
The Chinese University of Hong Kong  
Hong Kong SAR, China

Paul R. Langlais PhD  
Assistant Professor  
Department of Biochemistry and Molecular Biology  
College of Medicine  
Mayo Clinic in Arizona  
Scottsdale, AZ, USA

Harold E. Lebovitz MD, FACE  
Professor of Medicine  
State University of New York Health Science Center at Brooklyn  
Staten Island, NY, USA
Contributors

Pierre J. Lefèbvre MD, PhD, FRCP, MAE
Emeritus Professor of Medicine
University of Liège;
Consultant Physician at University Hospital
CHU Sart Tilman
Liège, Belgium

Derek LeRoith MD, PhD
Professor of Medicine and
Director of Research
Division of Endocrinology, Diabetes & Bone Diseases
Icahn School of Medicine at Mount Sinai (ISMMS)
New York, NY, USA

José e Leroux-Stewart MD
Division of Endocrinology
Centre hospitalier de l’Université de Montréal
Institut de Recherches Cliniques de Montréal (IRCM)
Montréal, QC, Canada

Romina Lomonaco MD
Assistant Professor
Division of Endocrinology, Diabetes & Metabolism
University of Florida
Gainesville, FL, USA

Richard J. MacIsaac Bsc (Hons), PhD, MBBS, FRACP
Professor and Director
Department of Endocrinology & Diabetes
St Vincent’s Hospital & University of Melbourne
Melbourne, VIC, Australia

Dianna J. Magliano BAppSci (Hons), MPH, PhD
Associate Professor and Senior Epidemiologist
Baker IDI Heart and Diabetes Institute
Melbourne, VIC, Australia

Rayaz Malik BSc (Hons), MBChB, PhD, FRCP
Professor of Medicine
Weil Cornell Medical College
Qatar Foundation
Education City
Doha, Qatar

Lawrence J. Mandarino PhD
Division of Endocrinology
Mayo Clinic in Arizona
Scottsdale, AZ;
Center for Metabolic and Vascular Biology
Arizona State University
Tempe, AZ, USA

Jim I. Mann CNZM, MA, DM, PhD, FRACP, FRSNZ
Professor in Human Nutrition and Medicine
Department of Human Nutrition;
Director, Edgar Diabetes and Obesity Research Centre
University of Otago
Dunedin, New Zealand

Chinmay S. Marathe MBBS
Advanced Trainee in Endocrinology
Royal Adelaide Hospital
Research Fellow
Centre of Research Excellence (CRE) in Translating Nutritional Science to Good Health
Discipline of Medicine, University of Adelaide
Adelaide, SA, Australia

Ashley Marchante BS
Graduate Research Assistant
Department of Pediatrics
University of Miami Miller School of Medicine
Miami, FL, USA

Piero Marchetti
Associate Professor of Endocrinology
Department of Endocrinology
University of Pisa
Pisa, Italy

Andrea Mari BEng
Professor of Internal Medicine
Institute of Neuroscience
National Research Council
Padova, Italy

Massimo Massi Benedetti MD
President and Scientific Director
Hub for International Health ReSearch HIRS
IDF Senior Programme Adviser
Perugia, Italy

Teresa L. Mastracci PhD
Research Assistant Professor
Department of Pediatrics
Indiana University School of Medicine
Indianapolis, IN, USA

Karl Matz
Chief Physician
Danube University of Krems
Krems, Austria

Donald A. McClain MD, PhD
Department of Medicine
University of Utah and Veterans Affairs Medical Center
Salt Lake City, UT, USA

Juris J. Meier MD, FRCP Edin.
Division of Diabetes and Gastrointestinal Endocrinology
University Hospital St. Josef-Hospital
Ruhr-University Bochum Gudrunstr.
Bochum, Germany
Arne Melander MD, PhD
Professor Emeritus of Clinical Pharmacology
Lund University
Lund and Malmö, Sweden

Kohtaro Minami PhD
Associate Professor
Division of Cellular and Molecular Medicine
Kobe University Graduate School of Medicine
Kobe, Japan

Raghavendra G. Mirmira MD, PhD
Lilly Professor of Pediatric Diabetes
Professor of Pediatrics, Medicine, Physiology, and Biochemistry
Director, Diabetes Research Center
Indianna University School of Medicine
Indianapolis, IN, USA

M. Mojaddidi MBChB, PhD
Dean of Taibah University Research & Consulting Institute
Taibah University
Al-Madinah Al-Munawarah
Saudi Arabia

Sunder Mudaliar MD, FRCP (Edin), FACP, FACE
Clinical Professor of Medicine
University of California at San Diego;
Staff Physician
Section of Diabetes/Metabolism
VA San Diego Healthcare System
San Diego, CA, USA

Debabrata Mukherjee, MD, FACC, FAHA, FSCAI
Chairman, Department of Internal Medicine;
Chief, Cardiovascular Medicine
Professor of Internal Medicine
Texas Tech University
El Paso, TX, USA

Medha Munshi MD
Director of the Geriatric Diabetes Clinic
Joslin Diabetes Center
Beth Israel Deaconess Medical Center
Department of Medicine
Harvard Medical School
Boston, MA, USA

Mary Beth Murphy RN, MS, CDE, MBA, CCRP
Research Nurse Director
The University of Tennessee Health Science Center
Memphis, TN, USA

K.M. Venkat Narayan
Centers for Disease Control and Prevention
National Center for Chronic Disease Prevention and Health Promotion
Atlanta, GA, USA

Andrea Natali MD
Professor of Internal Medicine
Department of Clinical and Experimental Medicine
Section of Internal Medicine
University of Pisa
Pisa, Italy

Michael A. Nauck MD
Professor of Internal Medicine
Division of Diabetology and Gastrointestinal Endocrinology
Medical Department 1
St. Josef-Hospital
Ruhr-University Bochum
Bochum, Germany

Karl J. Neff MB, MRCPI
Clinical Research Fellow
Diabetes Complications Research Centre
Conway Institute of Biomolecular and Biomedical Research
University College Dublin
Dublin, Ireland

John W. Newman PhD
Research Chemist, Obesity & Metabolism Research Unit
United States Department of Agriculture-Agricultural Research Service
Western Human Nutrition Research Center—and—
Adjunct Associate Professor, Department of Nutrition
University of California
Davis, CA, USA

Ruslan Novosyadlyy MD, PhD
VP
Department of Oncology — Translational Medicine
ImClone Systems, a wholly-owned subsidiary of Eli Lilly and Company
Alexandria Center for Life Science
New York, NY, USA

Kwame Ntim MD
Assistant Professor
Division of Endocrinology, Diabetes & Metabolism
University of Florida
Gainesville, FL, USA

Miladys M. Palau-Collazo MD
Pediatric Endocrine Fellow
Department of Pediatrics
Yale University School of Medicine
New Haven, CT, USA

Sianna Panagiotopoulos PhD
Manager of Research
Endocrine Centre & Department of Medicine
Austin Health;
University of Melbourne
Melbourne, VIC, Australia
Mie Jung Park-York PhD
Research Assistant Professor
Department of Biology
Utah State University
Logan, UT, USA

Liza Phillips MBBS (Hons), PhD, MS, FRACP
Postdoctoral Research Fellow
Royal Adelaide Hospital / Centre of Research Excellence (CRE) in Translating Nutritional Science to Good Health
Discipline of Medicine, University of Adelaide
Adelaide, SA, Australia

Jorge Plutzky MD
Director, The Vascular Disease Prevention Program
Preventive Cardiology, Cardiovascular Division
Brigham and Women’s Hospital
Boston, MA, USA

Vincent Poitout DVM, PhD, FCAHS
Professor of Medicine
Montreal Diabetes Research Center, CRCHUM;
Departments of Medicine and Biochemistry
Université de Montréal
Montréal, QC, Canada

David Preiss MD, PhD
Clinical Senior Lecturer
Institute of Cardiovascular and Medical Sciences
BHF Glasgow Cardiovascular Research Centre
University of Glasgow
Glasgow, UK

Erosha Premaratne MBBS, MD, FRACP
Endocrinologist
Endocrine Centre & Department of Medicine
Austin Health;
University of Melbourne
Melbourne, VIC, Australia

Alberto Pugliese MD
Professor of Medicine, Immunology and Microbiology
Division of Diabetes, Endocrinology and Metabolism
Head, Immunogenetics Program
Diabetes Research Institute
University of Miami Miller School of Medicine
Miami, FL, USA

Qing Qiao
Adjunct Professor
Department of Public Health
University of Helsinki;
National Public Health Institute
Helsinki, Finland

Rémi Rabasa-Lhoret MD, PhD
Associate Professor
Division of Endocrinology
Centre hospitalier de l’Université de Montréal (CHUM);
Department of Nutrition
Université de Montréal
Institut de Recherches Cliniques de Montréal (IRCM)
Montréal, QC, Canada

Rajesh Rajendran MBBS, AHEA, MRCP(UK)
Specialist Registrar in Diabetes and Endocrinology
Academic Department of Diabetes and Endocrinology
Queen Alexandra Hospitals
Portsmouth NHS Trust
Portsmouth, UK

Daiva Rastenyte
Professor
Lithuanian University of Health Sciences
Kaunas, Lithuania

Eric Ravussin PhD
Professor and Douglas L. Gordon Chair in Diabetes & Metabolism
Associate Executive Director, Clinical Sciences
Director, Nutrition Obesity Research Center
Pennington Biomedical Research Center
Louisiana State University
Baton Rouge, LA, USA

Simon D. Rees BSc, PhD
Postdoctoral Research Fellow
Clinical and Experimental Medicine
College of Medicine and Dentistry
University of Birmingham;
Diabetes Centre
Heart of England NHS Foundation Trust
Birmingham, UK

Marian J. Rewers MD, PhD
Professor of Pediatrics & Medicine
Executive Director
Barbara Davis Center for Childhood Diabetes
University of Colorado Denver
Aurora, CO, USA

Gissette Reyes-Soffer MD
Assistant Professor of Medicine
Division of Preventive Medicine and Nutrition
Columbia University College of Physicians and Surgeons
New York, NY, USA

Christopher J. Rhodes BSc (Hons), PhD
Professor, Research Director and Chair of Committee on Molecular Metabolism
Kovler Diabetes Center
University of Chicago
Chicago, IL, USA
Erik A. Richter MD, DM Sci  
Professor of Human Physiology and Exercise Physiology  
Molecular Physiology Group  
Department of Nutrition, Exercise and Sports  
Faculty of Science, University of Copenhagen  
Copenhagen, Denmark

R. Paul Robertson MD  
Professor of Medicine and Pharmacology  
University of Washington;  
Professor of Medicine  
University of Minnesota;  
President Emeritus  
Pacific Northwest Diabetes Institute  
Seattle, WA, USA

Arlan L. Rosenbloom MD  
Adjunct Distinguished Service Professor Emeritus  
University of Florida College of Medicine  
Department of Pediatrics  
Gainesville, FL, USA

Carel W. le Roux MBChB, FRCP, FRCPath, PhD  
Diabetes Complications Research Centre  
Conway Institute  
University College Dublin  
Dublin, Ireland

Jorge Ruiz MD  
Associate Director for Education and Evaluation  
GRECC-Miami VAHS  
Division of Geriatric Medicine, Departments of Medicine  
University of Miami Miller School of Medicine  
Miami, FL, USA

Susan L. Samson MD, PhD, FRCP, FACE  
Assistant Professor  
Department of Medicine  
Baylor College of Medicine  
Houston, TX, USA

Naveed Sattar MBChB, PhD, FRCPath (Glas), FRCP  
Professor of Metabolic Medicine  
Institute of Cardiovascular and Medical Sciences  
BHF Glasgow Cardiovascular Research Centre  
University of Glasgow  
Glasgow, UK

Oliver Schnell MD, PhD  
Executive Member of the Managing Board of the Forschergruppe Diabetes e.V. at the Helmholtz Center  
Munich  
Munich — Neuherberg, Germany

Susumu Seino MD, DM Sci  
Professor and Head  
Division of Molecular and Metabolic Medicine  
Department of Physiology and Cell Biology  
Kobe University Graduate School of Medicine  
Kobe, Japan

A.K. Sharma BVSc., PhD  
Professor of Anatomy  
Uttaranchal Ayurvedic Medical College  
Dehradun, India

Jonathan E. Shaw MD, FRACP, FRCP (UK)  
Associate Director and Associate Professor  
Baker IDI Heart and Diabetes Institute  
Melbourne, VIC, Australia

Tadao Shibasaki PhD  
Assistant Professor  
Division of Cellular and Molecular Medicine  
Kobe University Graduate School of Medicine  
Kobe, Japan

Jay S. Skyler MD, MACP  
Professor of Medicine, Pediatrics, & Psychology  
Division of Endocrinology, Diabetes, & Metabolism  
Deputy Director - Diabetes Research Institute  
University of Miami Miller School of Medicine  
Miami, FL, USA

Eberhard Standl MD, PhD  
Professor of Medicine  
Munich Diabetes Research Group e.V.  
Helmholtz Centre  
Munich, Germany

Andrea K. Steck MD  
Assistant Professor of Pediatrics  
Barbara Davis Center for Childhood Diabetes  
University of Colorado Denver  
Aurora, CO, USA

Roland Stein PhD  
Professor  
Vanderbilt University Medical Center  
Nashville, TN, USA

Douglas F. Stickle PhD, DABCC, FACB  
Professor, Department of Pathology  
Director, Chemistry and POCT  
Jefferson Medical College, Thomas Jefferson University  
Philadelphia, PA, USA
Contributors

Bernd Stratmann PhD
Research Director Diabetes Center
Heart and Diabetes Center NRW
University Clinic Ruhr University
Bochum, Germany

Abd A. Tahrani MD, MRCP, MMedSci, PhD
NIHR Clinician Scientist
Head of Weight Management Services
Honorary Consultant in Endocrinology and Diabetes
Centre of Endocrinology, Diabetes and Metabolism (CEDAM)
School of Clinical and Experimental Medicine
University of Birmingham
Birmingham, UK

Charmaine S. Tam, PhD
Postdoctoral Fellow
Pennington Biomedical Research Center
Louisiana State University
Baton Rouge, LA, USA;
The Charles Perkins Centre and School of Biological Sciences,
University of Sydney
Sydney, NSW, Australia

William V. Tamborlane MD
Professor and Chief of Pediatric Endocrinology
Yale School of Medicine and the Yale Center for Clinical Investigation
New Haven, CT, USA

Gavin S. Tan MBBS (S’pore), MMed (Ophth), FRCSEd (Ophth), FAMS
Consultant & Assistant Professor
Vitreo-Retinal Service
Singapore National Eye Centre;
Singapore Eye Research Institute
Duke-NUS Graduate Medical School
Singapore

Yoel Toledano MD
Division of Maternal Fetal Medicine
Helen Schneider Hospital for Women
Rabin Medical Center
Petah-Tiqva, Israel

Dace L. Trence MD
Professor of Medicine
University of Washington Medical Center
Seattle, WA, USA

Diethelm Tschoepe MD, FESC
Professor of Medicine
Medical Director Diabetes Center
Heart and Diabetes Center NRW
University Clinic Ruhr University
Bochum, Germany

Jaakko Tuomilehto MD, MA (sociol), PhD, FRCP (Edin), FESC
Professor
Department of Public Health
University of Helsinki;
National Public Health Institute
Helsinki, Finland;
Danube University of Krems
Krems, Austria;
King Abdulazziz University
Jeddah, Saudi Arabia

Guillermo E. Umpierrez MD
Professor of Medicine
Department of Medicine, Division of Endocrinology and Metabolism
Emory University School of Medicine
Chief of Diabetes and Endocrinology,
Grady Health System
Atlanta, GA, USA

Luis Varela
Program in Integrative Cell Signalling and Neurobiology of Metabolism
Section of Comparative Medicine
Yale University School of Medicine
New Haven, CT, USA

Elena Volpi, MD, PhD
Daisy Emery Allen Distinguished Chair in Geriatric Medicine
Director ad interim, Sealy Center on Aging
Associate Director, Institute for Translational Sciences
University of Texas Medical Branch
Galveston, TX, USA

Paul Welsh PhD
British Heart Foundation Research Fellow
Institute of Cardiovascular and Medical Sciences
BHF Glasgow Cardiovascular Research Centre
University of Glasgow
Glasgow, UK

Desmond E. Williams
Centers for Disease Control and Prevention
National Center for Chronic Disease Prevention and Health Promotion
Atlanta, GA, USA

Rhys Williams MA, PhD, FFPH, FRCP
Emeritus Professor of Clinical Epidemiology
Swansea University
Swansea, UK

Tien Yin Wong MBBS, MMED (Ophth), MPH, PhD, FRCSE, FRANZCO, FAMS
Provost Chair Professor of Ophthalmology
Singapore Eye Research Institute, Singapore National Eye Centre
Chairman, Department of Ophthalmology
National University of Singapore & National University Hospital
Vice-Dean, Office of Clinical Sciences
Duke NUS Graduate Medical School
Singapore
Thomas Yates PhD, MSc, BSc  
Reader in Physical Activity, Sedentary Behaviour and Health  
Leicester Diabetes Centre  
College of Medicine, Biological Sciences and Psychology  
University of Leicester;  
National Institute for Health Research Leicester-Loughborough Diet, Lifestyle and Physical Activity  
Biomedical Research Unit  
Leicester, UK  

Hannele Yki-Järvinen MD, FRCP  
Professor of Medicine  
University of Helsinki  
Minerva Foundation Institute for Medical Research  
Helsinki, Finland  

David A. York PhD  
Professor  
Department of Biology  
Utah State University  
Logan, UT, USA  

Barak Zafrir, MD  
Preventive Cardiology, Cardiovascular Division  
Brigham and Women’s Hospital  
Boston, MA, USA;  
Department of Cardiovascular Medicine  
Lady Davis Carmel Medical Center  
Haifa, Israel  

Paul Zimmet AO, MD, PhD, FRACP, FRCP, FTSE  
Director Emeritus, Victor Smorgon Diabetes Centre  
Baker IDI Heart and Diabetes Institute  
Adjunct Professor, Monash University  
Honorary President, International Diabetes Federation  
Chair, Programme Committee, World Diabetes Congress 2013  
Melbourne, VIC, Australia
As the epidemic of diabetes continues to expand in parallel with the rapid spread of obesity, healthcare providers strive to find interventions to reduce the morbidity, mortality, and rising costs associated with this devastating disease, which ravages both the micro- and the macrovasculature. Although the increase in incidence of type 2 diabetes may be attributed to the expanding girth of the population coupled with a lack of physical activity, the marked increase in the incidence of type 1 diabetes remains unexplained. Our knowledge of the cellular, biochemical, and molecular etiology of impaired insulin action and beta-cell failure has expanded enormously, but the genetic basis of both type 1 and type 2 diabetes and their associated complications is still by and large undefined. Despite the introduction of multiple new classes of antidiabetic agents for the treatment of type 2 diabetes, and newer insulin preparations, insulin delivery systems, and glucose-sensing devices for the management of type 1 diabetes, glycemic control is suboptimal in approximately half of all diabetic patients and the excess risk for macrovascular complications is largely unexplained. In many parts of the world, these treatment advances are not available and instituting behavioral modification programs at the societal and individual level is proving to be inadequate in curbing the growing epidemic of obesity. Whether the introduction of novel weight loss medications will stem the tide of obesity remains to be determined.

The fourth edition of the International Textbook of Diabetes Mellitus will continue to be the most widely referenced textbook of diabetes worldwide and draws upon the expertise of leading basic scientists, clinicians, educators, and healthcare professionals globally to provide the most updated information on advances in diabetes research and clinical care. This information will be an invaluable resource and provide the practicing physician, as well as the basic scientist and clinical investigator, with the requisite resources to advance them to the frontiers of biomedical research in the fields of diabetes, metabolism, and obesity and to provide them with state-of-the-art knowledge to optimize clinical care for their diabetic patients.

Ralph A. DeFronzo, MD
SECTION I

Epidemiology
CHAPTER 1

Classification of diabetes mellitus and other categories of glucose intolerance

Dianna J. Magliano, Paul Zimmet and Jonathan E. Shaw
Baker IDI Heart and Diabetes Institute, Melbourne, VIC, Australia

Key points
- The classification and diagnosis of diabetes is based on etiology and not on pharmacologic treatment.
- Diagnoses of diabetes are made using fasting plasma glucose, 2-hour postchallenge of glucose or HbA1c.
- Differentiation between type 1 and type 2 diabetes is usually straightforward but can be difficult among obese children and adults.
- Precise diagnoses of certain monogenic diabetes using genetic testing can be useful as the outcomes can influence treatment decisions.
- A range of commonly used drugs such as statins and glucocorticoid steroids can lead to the development of diabetes.

Introduction

A critical requirement for orderly epidemiologic, genetic and clinical research, and indeed for the management of diabetes mellitus and other forms of glucose intolerance is an appropriate classification system. Furthermore, a hallmark in the process of understanding the etiology of a disease and studying its natural history is the ability to identify and differentiate its various forms and place them into a rational etiopathologic framework. While there have been a number of sets of nomenclature and diagnostic criteria proposed for diabetes, no systematic categorization existed until the mid 1960s [1]. Now diabetes mellitus is recognized as being a syndrome, a collection of disorders that have hyperglycemia and glucose intolerance as their hallmark, due either to insulin deficiency or to impaired effectiveness of insulin’s action, or to a combination of these.

Historical perspective and current classifications

Previous classifications
In 1965, an Expert Committee on Diabetes Mellitus published the first World Health Organization (WHO) report on diabetes classification [1]. The report includes one of the first attempts at international consensus on a classification. They decided to classify diabetes: “… based on the age of recognized onset, which seemed to be the only reliable means of classification for universal use.”

The report also recognized certain specific types of diabetes including brittle, insulin-resistant, gestational, pancreatic, endocrine, and iatrogenic diabetes. Since then, several pathogenic mechanisms have been described and long-term studies have shown different courses and outcomes of different types of diabetes.

A revised classification of glucose intolerance, was formulated by the National Diabetes Data Group (NDDG) [2]. This was amended and adopted in the second report of the WHO Expert Committee in 1980 [3] and in a modified form in 1985. The 1980 Expert Committee proposed two major classes of diabetes mellitus and named them insulin-dependent diabetes mellitus (IDDM) or type 1, and non-insulin-dependent diabetes mellitus (NIDDM) or type 2 [3]. In the 1985 Study Group Report, the terms type 1 and type 2 were omitted, but the classes IDDM and NIDDM were retained and a new class of malnutrition-related diabetes mellitus (MRDM) was introduced [4]. The 1985 WHO classification was essentially based on clinical descriptions, with a specific focus on the pharmacologic management of patients (i.e., insulin-dependent, non-insulin-dependent, gestational). The question as to whether certain clinical forms...
of diabetes (such as the so-called “tropical diabetes”) had been given adequate priority to correct hierarchic order that was raised many years before probably led to the introduction of MRDM, although more precise epidemiologic data and a better assessment were needed, and called for.

Both the 1980 and 1985 reports included other types of diabetes and impaired glucose tolerance (IGT) as well as gestational diabetes mellitus (GDM). The 1985 classification was widely accepted and used internationally, and represented a compromise between clinical and etiological classifications. Furthermore, it permitted classification of individual patients in a clinically useful manner even when the specific etiology was unknown. The 2011 American Diabetes Association (ADA) [5] classifications or staging of diabetes still include clinical descriptive criteria but a complementary classification according to etiology is recommended by both organizations.

In 1999, the WHO incorporated an approach developed by Kuzuya and Matsuda [6], which clearly separated the criteria related to etiology from those related to the degree of deficiency of insulin or insulin action, and defined each patient on the basis of these two sets of criteria (Figure 1.1). It is now well established that diabetes may progress through several clinical stages during its natural history, quite independent of its etiology. The clinical staging reflects this and, indeed, individuals may move from one stage to another stage in both directions (Figure 1.1). Even if there is no information concerning the underlying etiology, persons with diabetes or those who are developing the disease can be categorized by stage according to clinical characteristics.

**Current classification**

The current classification allows for various degrees of hyperglycemia in individuals irrespective of the disease process. These are glycemic stages ranging from normoglycemia (normal glucose tolerance) to hyperglycemia where insulin is required for survival. All individuals with the disease can be categorized according to clinical stage [7]. The stage of glycemia may change over time depending on the extent of the underlying disease processes. As shown in Figure 1.1, the disease process may be present but may not have progressed far enough to cause hyperglycemia. The etiological classification is possible as the defect or process which may lead to diabetes may be identified at any stage in the development of diabetes, even at the stage of normoglycemia. As an example, the presence of islet cell antibodies (ICA) and/or antibodies to glutamic acid decarboxylase (anti-GAD) [8] in a normoglycemic individual indicates the autoimmune process, which underlies type 1 diabetes, is present, although the individual may or may not ultimately develop diabetes [7,9]. For type 2 diabetes, there are few useful highly specific indicators, though the presence of risk factors such as obesity indicates the likelihood of developing type 2 diabetes. Hopefully, future research will reveal some specific markers of the type 2 diabetes disease process.

<table>
<thead>
<tr>
<th>Types</th>
<th>Stages</th>
<th>Normoglycemia: Normal glucose tolerance</th>
<th>Hyperglycemia</th>
<th>Diabetes Mellitus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>IGT and/or IFG</td>
<td>Not insulin requiring</td>
</tr>
<tr>
<td>Type 1</td>
<td>• Autoimmune</td>
<td>• Idiopathic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type 2</td>
<td>• Predom.</td>
<td>• insulin resistance</td>
<td>• Predom. insulin secretory defects</td>
<td></td>
</tr>
<tr>
<td>Other specific types</td>
<td>• Genetic defects of β-cell function</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Genetic defects of insulin action</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Diseases of exocrine pancreas</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Endocrinopathies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Drug or chemical induced</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Others</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gestational hyperglycemia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 1.1** Disorders of glycemia: etiologic types and clinical stages. Source: World Health Organization 1999 [7]. Reproduced with permission of the WHO.
The same disease process can cause various degrees of impaired glucose metabolism such as impaired fasting glycemia (IFG) and impaired glucose tolerance (IGT) without fulfilling the criteria for the diagnosis of diabetes [7]. Weight reduction, exercise and/or oral hypoglycemic therapy can achieve satisfactory glycemic control in some persons with type 2 diabetes. These persons, therefore, do not require insulin initially but may do so much later in their course as β-cell function deteriorates. Some persons require insulin for adequate glycemc control at an earlier stage in type 2 diabetes but could survive without it. By definition these persons have some residual insulin secretion. Patients with extensive β-cell destruction (minimal residual insulin secretion) do require insulin for survival and this is the hallmark of type 1 diabetes [7,9].

The classification by etiological type (Table 1.1) results from improved understanding of the causes of diabetes, although this is still far from complete, particularly for type 1 diabetes.

The terms “insulin-dependent diabetes mellitus,” “non-insulin-dependent diabetes mellitus” and their acronyms “IDDM” and “NIDDM” have been removed from classifications. These terms were very confusing and frequently resulted in misclassification, as patients were classified on the basis of their treatment, and indeed their age, rather than on pathogenesis. In the current classification, the terms “type 1” and “type 2” are retained (using Arabic rather than Roman numerals) [7].

Type 1 includes those cases attributable to an autoimmune process (although the basic precipitating cause of this process is still unknown), as well as those with β-cell destruction for which neither an etiology nor a pathogenesis is known (idiopathic). Those forms of β-cell destruction or failure to which specific causes can be assigned (e.g. cystic fibrosis, mitochondrial defects) are not included in this type of diabetes. These issues are discussed in greater detail later.

Type 2 includes the common major form of diabetes which results from defect(s) in insulin secretion and/or from insulin resistance, and often a combination of both. Malnutrition-related diabetes (MRDM) is no longer part of the WHO classification [7]. Of its two subtypes, protein-deficient pancreatic diabetes (PDPD or PDDM) needs more studies for a better definition. The other former subtype of MRDM, fibrocalculus pancreatic diabetes (FCPD), is now classified as a disease of the exocrine pancreas labeled “fibrocalculus pancreatopathy”, which may lead to diabetes.

Impaired glucose tolerance (IGT) and impaired fasting glycemia (IFG) are classified as stages of impaired glucose regulation, since they can be observed in any hyperglycemic disorder.

Gestational diabetes is a state of glucose intolerance first recognized during pregnancy which usually resolves after delivery but is associated with later increased long-term risk of type 2 diabetes. It encompasses the groups formerly classified as gestational impaired glucose tolerance (GIGT) and gestational diabetes mellitus (GDM) [7].

### DIABETES TYPES

#### Type 1 process

Type 1 indicates the processes of β-cell destruction that may ultimately lead to diabetes in which insulin is required for survival in order to prevent the development of ketoacidosis, coma, and death. This category comprises:

- **Immune-mediated diabetes mellitus:** This is the classical form of type 1 diabetes, which can occur at any age, and results from a cell-mediated autoimmune destruction of the pancreatic β cells. The type 1 process is characterized by the presence of ICA, anti-GAD, islet antigen 2 (IA2) or insulin autoantibodies which identify the autoimmune process associated with β-cell destruction [9]. Other autoimmune disorders such as Grave’s disease, Hashimoto’s thyroiditis and Addison’s disease may be associated with type 1 diabetes mellitus [9].

The rate of β-cell destruction is quite variable, typically being rapid in children and slower in adults. Typically, type 1 diabetes requires insulin therapy from the time of presentation in both adults and children, but a slowly progressive form, latent autoimmune diabetes in adults (LADA), is well described [8]. Blood glucose in LADA can initially be controlled by lifestyle change and oral hypoglycemic agents, and may therefore masquerade as type 2 diabetes. However, in comparison to the typical patient with type 2 diabetes, LADA patients are leaner and
progress much more rapidly to requiring insulin. Importantly, markers of autoimmunity (most commonly anti-GAD antibodies) are present, and therefore LADA falls within type 1 autoimmune diabetes.

- **Idiopathic:** There are some forms of type 1 diabetes which have no known etiology, and no evidence of autoimmunity. Some of these patients have permanent insulinopenia and are prone to ketoacidosis [10]. This form is more common among individuals of African and Asian origin [11].

### Type 2 process

Type 2 diabetes is the commonest form of diabetes and is characterized by disorders of insulin resistance and insulin secretion, either of which may be the predominant feature. Both are usually present at the time when diabetes is clinically manifest. Insulin levels may be normal or even elevated at the time when diabetes is diagnosed. However, in the setting of insulin resistance, these levels are inadequate to maintain normoglycemia. This relative insulin deficiency is what differentiates diabetic insulin-resistant individuals from normoglycemic insulin-resistant individuals. Indeed, it is noteworthy that, to date, the majority of the genes that have been associated with type 2 diabetes are related to insulin secretion, and not to insulin resistance [12].

At least initially, and often throughout their lifetime, these individuals do not need insulin treatment to survive [13]. Type 2 diabetes is frequently asymptomatic and undiagnosed for many years because the hyperglycemia is often not severe enough to provoke noticeable symptoms [14]. Nevertheless, such patients are at increased risk of developing macrovascular and microvascular complications. Type 2 diabetes is a very heterogeneous disorder and there are certainly many different causes of this form of diabetes. However, it is likely that the number of patients placed in this category will decrease in the future as identification of specific pathogenic processes and genetic defects permit better differentiation and a more definitive classification. Although the specific etiologies of type 2 diabetes are not known, autoimmune destruction of the pancreas does not occur and patients do not have any of the other specific causes of diabetes listed in Table 1.2.

Most patients with the type 2 process of diabetes are overweight or obese, and obesity itself causes insulin resistance. Many of those not obese by traditional criteria, for example body mass index, may have an increased percentage of body fat distributed predominantly in the abdominal region [13]. Ketoacidosis seldom occurs in type 2 diabetes and when seen, it usually arises in association with the stress of another illness such as infection. Ketosis-prone atypical diabetes, also referred to as ketosis-prone type 2 diabetes is characterized by presentation with severe hyperglycemia and ketoacidosis requiring immediate insulin therapy [15]. More than 50% of these individuals will revert to an insulin-free near-normoglycemia

### Table 1.2 Other specific types of diabetes [7]

<table>
<thead>
<tr>
<th>Genetic defects of β-cell function</th>
<th>HNF1A MODY</th>
<th>HNF4A MODY</th>
<th>HNF1B MODY</th>
<th>GCK MODY</th>
<th>MTDNA 3243 MIDD</th>
<th>KCNJ11 PNDM</th>
<th>KCNJ11 DEND</th>
<th>6q24 TNDM</th>
<th>ABCG8 TNDM</th>
<th>INS PNDM</th>
<th>WFS1 Wolfram syndrome</th>
<th>FOXP3 IPEX syndrome</th>
<th>EIF2AK3 Wolcott–Rallison syndrome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genetic defects in insulin action</td>
<td>INSR Type A insulin resistance</td>
<td>INSR Leprechaunism</td>
<td>INSR Rabson–Mendenhall syndrome</td>
<td>LMNA FPLD</td>
<td>PPPARG FPLD</td>
<td>AGR1 P2 CGL</td>
<td>BSCL CGL</td>
<td>Lipotropic diabetes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diseases of the exocrine pancreas</td>
<td>Fibrocalculous pancreatopathy</td>
<td>Pancreatitis</td>
<td>Trauma / pancreatectomy</td>
<td>Neoplasia</td>
<td>Cystic fibrosis</td>
<td>Hemochromatosis</td>
<td>Others</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Endocrinopathies</td>
<td>Cushing syndrome</td>
<td>Aldosteronoma</td>
<td>Acromegaly</td>
<td>Pheochromocytoma</td>
<td>Glucagonoma</td>
<td>Hyperthyroidism</td>
<td>Somatostatinoma</td>
<td>Others</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drug- or chemical-induced (see Table 1.3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infections</td>
<td>Congenital rubella</td>
<td>Cytomegalovirus</td>
<td>Others</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uncommon forms of immune-mediated diabetes</td>
<td>Insulin autoimmune syndrome (antibodies to insulin)</td>
<td>Anti–insulin receptor antibodies</td>
<td>“Stiff man” syndrome</td>
<td>Others</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other genetic syndromes (see Table 1.4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:** Nomenclature: the gene name is followed by the clinical syndrome with the gene number designated using the HUGO convention.

- MODY maturity onset diabetes of the young
- MIDD maternally inherited diabetes and deafness
- PNDM permanent neonatal diabetes mellitus
- DEND development delay epilepsy
- TNDM transient neonatal diabetes mellitus

Source: World Health Organization 1999 [7]. Reproduced with permission of the WHO.
within weeks or months with multiorgan insulin resistance not dissimilar to type 2 diabetes [16]. This condition is commonly found in sub-Saharan Africa and African migrants and is referred to as “Flatbush diabetes” [17].

The risk of developing type 2 diabetes increases with age, obesity, and lack of physical activity. It occurs more frequently in women with prior GDM, in those with hypertension or dyslipidemia, and its frequency varies between different ethnic subgroups [7]. Type 2 diabetes is often associated with strong familial, likely genetic, predisposition but the genetics of type 2 diabetes are quite complex and not clearly defined [18]. Some patients who present a clinical picture consistent with type 2 diabetes have been shown to have antibodies similar to those found in type 1 diabetes.

Although diagnosis in most patients with type 2 diabetes is made in adult years, the disease is now increasingly seen in adolescents and even children, especially in a background of high obesity prevalence. At presentation, ketosis or even ketoacidosis, may occur in this younger age group and insulin is often required in the initial management. However, once the acute metabolic disturbance is rectified, insulin can often be withdrawn, and glycemic control achieved with lifestyle measures and oral pharmacotherapy.

Other specific types

The other specific types of diabetes are less common and can be broadly classed as genetic, exocrine pancreatic, endocrine, and drug-induced causes [7]. A more comprehensive breakdown is provided in Table 1.2 and the more common types are discussed briefly later.

Classification of genetic disorders

With ongoing advances in the study of molecular genetics, there has been considerable progress in the identification of specific subtypes of diabetes of genetic origin. Through this work, it has been shown that the clinical subgroups are heterogeneous and there has been recognition of several novel, genetic-based syndromes associated with diabetes. The progress in our ability to examine genes to arrive at a diabetes diagnosis has improved treatment for these patients [19] and thus genetic diagnosis has become a key part of clinical management in many countries.

Genetic defects of β-cell function

The diabetic state may be associated with monogenic defects in β-cell function. These forms are characterized by onset of mild hyperglycemia during childhood or early adulthood, and include maturity-onset diabetes of the young (MODY), permanent neonatal diabetes (PNDM), transient neonatal diabetes (TNDM), and many other insulin-deficient syndromes with a myriad of other clinical features [7]. The most well characterized of these is MODY. MODY is inherited in an autosomal dominant pattern and typically presents before the age of 25 years. While the condition results from β-cell dysfunction, it is not always insulin dependent. Molecular genetic testing can define a diagnosis in 1–2% of all diabetic patients with monogenic diabetes. Advances in this field have led to the identification of the genes associated with many clinically identified subgroups of diabetes and explained clinical heterogeneity in conditions defined by age of diagnosis, for example neonatal diabetes and MODY. Molecular genetic tests are now available to help define the diagnosis, and importantly alter prognosis and optimize treatment of children, young adults and their families with diabetes.

Several mutations associated with MODY have been identified to date, of which the most common genetic subtypes are: GCK MODY, HNF1A MODY, HNF4A MODY, and IPF1 MODY [19]. These are listed in Table 1.2. Among these subtypes, the HNF1A MODY subtype is the most common and results in a progressive and marked hyperglycemia with a high risk of microvascular and macrovascular complications [20], but these patients respond well to sulfonylureas [21]. Subtype HNF4A is similar to HNF1A but patients have marked macrosomia and transient neonatal hypoglycemia [22]. The other subtype, GCK MODY, is a milder form of diabetes, characterized by a mild fasting hyperglycemia that is generally lifelong with little deterioration with age and does not require treatment [23,24].

In children less than 6 months of age, diabetes is more likely to be monogenic than autoimmune type 1 diabetes [25]. However, in approximately 50% of these infants, the diabetes is transient (TNDM) [24]. Further to the specific genetic types mentioned here, there are also many subtypes of neonatal diabetes which present as a result of multisystem clinical syndromes [26]. For example, Wolfram syndrome, also referred to as DIDMOAD, is inherited by autosomal recessive trait, is a monogenic multisystem syndrome, and is characterized by marked β-cell dysfunction [27].

Point mutations in mitochondrial DNA have been found to be associated with diabetes and sensori-neural deafness [28] and lead to a condition known as maternally inherited diabetes and deafness (MIDD). Genetic abnormalities that result in the inability to convert proinsulin to insulin have been identified in a few families. Usually such traits are inherited in an autosomal dominant pattern [29] and the resultant carbohydrate intolerance is mild.

Genetic defects in insulin action

Genetic defects in insulin action are rare, and the associated metabolic abnormalities may range from hyperinsulinemia and modest hyperglycemia to severe symptomatic diabetes resulting in death [30]. Acanthosis nigricans may be present in some of these individuals. This syndrome was termed type A insulin resistance in the past. In such patients, diabetes only occurs when there is no β-cell response to the insulin resistance.
Two pediatric syndromes that have mutations in the insulin receptor gene with subsequent alterations in insulin receptor function and extreme insulin resistance are called Leprechaunism and the Rabson–Mendenhall syndrome [31]. A heterogeneous group of disorders of lipid storage characterized by lipodystrophy, in which insulin resistance is a common feature, has also been described [32].

**Diseases of the exocrine pancreas**

Pancreatitis, trauma, infection, pancreatic carcinoma, and pancreatectomy are some of the acquired processes of the pancreas that can cause diabetes. Any process that diffusely injures the pancreas may cause diabetes [33]. With the exception of cancer, damage to the pancreas must be extensive for diabetes to occur. However, adenocarcinomas that involve only a small portion of the pancreas have been associated with diabetes. This implies a mechanism other than a simple reduction in β-cell mass [34]. Hemochromatosis will also damage β cells and impair insulin secretion [35]. Fibrocalculous pancreatopathy may be accompanied by abdominal pain radiating to the back and pancreatic calcification on X-ray and ductal dilatation. Pancreatic fibrosis and calcified stones in the exocrine ducts are found at autopsy [36].

**Endocrinopathies**

Insulin action can be antagonized by several hormones (e.g. growth hormone, cortisol, glucagon, epinephrine). Diseases associated with excess secretion of these hormones can cause diabetes (e.g. acromegaly, Cushing syndrome, glucagonoma and pheochromocytoma) [7]. These forms of hyperglycemia resolve when the hormone excess is removed. Somatostatinoma and aldosteronoma-induced hypokalemia, can cause diabetes at least in part by inhibiting insulin secretion [37]. Hyperglycemia generally resolves following successful removal of the tumor.

**Drug-or chemical-induced diabetes**

Insulin secretion may be impaired by many drugs. They may not, by themselves, cause diabetes but may precipitate diabetes in persons with insulin resistance [38]. Pancreatic β-cell destruction may occur with the use of certain toxins such as Vacor (a rat poison) [39], pentamidine [40], and some immunosuppressive drugs. Among these β-cell toxic agents, the most commonly used are the immunosuppressive agents of which the calcineurin inhibitors (e.g. tacrolimus and cyclosporin) are the main culprits. While the main action of calcineurin inhibitors in inducing diabetes is by reducing insulin secretion by pancreatic β cells, these drugs may also increase insulin resistance [41]. There is good evidence to suggest that there is greater potential of tacrolimus to induce diabetes compared with cyclosporine [42]. Diabetes induced by these drugs may be permanent due to β-cell destruction, or may only occur while the drug is being taken, with recovery between treatment cycles [42].

Studies involving other immunosuppressive agents such as mycophenolate mofetil and sirolimus are few and results are inconsistent. Clinical studies have shown that daclizumab seems to have a neutral effect [43]. Patients receiving interferon alpha have been reported to develop diabetes associated with islet cell autoantibodies and, in certain instances, severe insulin deficiency [44].

There are also many drugs and hormones that can impair insulin action. The list shown in Table 1.3 is not all-inclusive, but reflects the more commonly recognized drug-, hormone-, or toxin-induced forms of diabetes and hyperglycemia. Among these, there are several commonly used diabetes-inducing drugs that deserve special mention. These include the HMG CoA reductase agents (statins), glucocorticoid steroids, anti-HIV agents and antipsychotic drugs.

**HMG CoA reductase agents**

HMG CoA reductase agents (statins) are commonly used drugs which have been purported to cause diabetes. Sattar et al. [45] reported that statin use compared to placebo increased risk of diabetes in a meta-analysis of 13 placebo-controlled trials. Another meta-analysis comparing intensive dose statin use with moderate statin therapy in five trials showed that the risk of developing diabetes was greater at higher statin doses [46]. The mechanism as to how statins cause diabetes is not known, but it has been suggested that these drugs may affect muscle and liver insulin sensitivity resulting in an increased diabetes risk [46]. It has also been suggested that the observed relationship between statins and diabetes is due to confounding as there is a tendency of individuals who take statins to have a high inherent risk of diabetes. Despite the increased risk of diabetes associated with statin use, a risk–benefit analysis has shown the beneficial nature of statins for cardiovascular disease (CVD), which outweighs the risk of diabetes associated with statin use [47].

**Antipsychotic agents**

There is accumulating evidence supporting an association of certain psychiatric conditions with type 2 diabetes which can be attributed to side-effects of treatment and a high baseline risk of diabetes in this patient group [48]. Diabetes

---

**Table 1.3 Drug or chemical-induced diabetes**

<table>
<thead>
<tr>
<th>Drug or Chemical</th>
<th>Diabetes Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nicotinic acid</td>
<td>Drug or chemical-induced</td>
</tr>
<tr>
<td>Glucocorticoids</td>
<td></td>
</tr>
<tr>
<td>Thyroid hormone</td>
<td></td>
</tr>
<tr>
<td>Alpha-adrenergic agonists</td>
<td></td>
</tr>
<tr>
<td>Beta-adrenergic agonists</td>
<td></td>
</tr>
<tr>
<td>Thiazides</td>
<td></td>
</tr>
<tr>
<td>Diltiazem</td>
<td></td>
</tr>
<tr>
<td>Pentamidine</td>
<td></td>
</tr>
<tr>
<td>Vacor</td>
<td></td>
</tr>
<tr>
<td>Interferon-alpha therapy</td>
<td></td>
</tr>
<tr>
<td>Statins</td>
<td></td>
</tr>
<tr>
<td>L-asparagine</td>
<td></td>
</tr>
<tr>
<td>Antipsychotic drugs, e.g. clozapine,</td>
<td></td>
</tr>
<tr>
<td>Highly active antiviral therapy, e.g. protease inhibitors</td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td></td>
</tr>
</tbody>
</table>
can be induced by the use of atypical antipsychotics including clozapine, olanzapine, risperidone, quetiapine, ziprasidone, and aripiprazole. These drugs have a direct effect of raising blood glucose and also lead to weight gain, [48] which subsequently may increase blood glucose levels.

Clozapine and olanzapine have been associated with a higher risk of diabetes than other antipsychotic agents in several studies [48]. These drugs have been associated with new-onset diabetes, exacerbation of pre-existing diabetes, and presentations with complications such as ketoacidosis. The data on risperidone and quetiapine in the studies mentioned earlier show inconsistent findings [48].

Atypical antipsychotics may have an independent effect on insulin sensitivity. Studies comparing insulin sensitivity in patients taking clozapine, olanzapine, or risperidone showed that those in clozapine and olanzapine groups had significantly decreased insulin sensitivity compared to risperidone groups. While there is generally less long-term data on aripiprazole and ziprasidone, a comparison of olanzapine and aripiprazole use in schizophrenic patients showed an increase in glucose in the olanzapine group [48].

Anti-HIV agents

Diabetes is fourfold more common in HIV-infected men exposed to highly active antiretroviral therapy (HAART) than HIV-negative men. Although most of the diabetes observed in this group is type 2 there has been a recent report of autoimmune diabetes and the development of anti-GAD antibodies after immune system recovery post HAART therapy [49], which suggests that type 1 diabetes can also arise in this group from treatment.

HAART is based on the use of a class of drugs known as protease inhibitors (PIs) and include atazanavir, darunavir, saquinavir, and ritonavir. PIs have been shown to increase insulin resistance and reduce insulin secretion, by interfering with GLUT-4 mediated glucose transport. PIs interfere with cellular retinoic acid-binding protein type 1 which interacts with peroxisomal proliferator-activated gamma (PPARγ) receptor. Inhibition of PPARγ promotes adipocyte inflammation, release of free fatty acids and insulin resistance [49]. Hyperglycemia resolves in almost all patients when PIs are discontinued [49] and all PIs do not have the same metabolic effects, with some drugs having a worse adverse effect than others.

Apart from HAART, another class of anti-HIV drugs associated with diabetes are the nucleoside analogs (reverse transcriptase inhibitors) (NRTIs) [50] especially when used for long periods of time [51]. The risk of diabetes is highest with stavudine, but the risk is also significant with zidovudine and didanosine. Proposed mechanisms include insulin resistance, lipodystrophy, and mitochondrial dysfunction [51]. It is postulated that PIs confer acute metabolic risks, while NRTIs confer cumulative risks of diabetes in predisposed, exposed persons. The use of both classes of drugs may be additive for diabetes risk [51].

**Glucocorticoids**

Glucocorticoids are the most common cause of drug-induced diabetes. They are used in the treatment of many medical conditions but are mostly prescribed for their anti-inflammatory effects [52]. They act through multiple pathways at the cellular and molecular levels, suppressing the cascades that would otherwise result in inflammation and promoting pathways that produce anti-inflammatory protein [53]. The mechanism by which glucocorticoids cause diabetes is thought to be mainly via insulin resistance, but there is also some evidence of effects on insulin secretion [54].

The effect of glucocorticoids is mainly on nonfasting glucose rather than fasting glucose levels [52], but there is uncertainty as to whether this reflects a relationship with clock time (perhaps linked to dosing times), or to a predominant effect on postprandial blood glucose levels.

**Infections**

Certain viruses have been associated with \(\beta\)-cell destruction. Diabetes occurs in some patients with congenital rubella [55]. Coxsackie B, cytomegalovirus, and other viruses (e.g. adenovirus and mumps) have been implicated in inducing diabetes [56–58].

**Uncommon but specific forms of immune-mediated diabetes mellitus**

Diabetes may be associated with several immunologic diseases with a pathogenesis or etiology different from that which leads to the type 1 diabetes process. Postprandial hyperglycemia of a severity sufficient to fulfill the criteria for diabetes has been reported in rare individuals who spontaneously develop insulin autoantibodies. However, these individuals generally present with symptoms of hypoglycemia rather than hyperglycemia [59]. The “stiff man syndrome” is an autoimmune disorder of the central nervous system, characterized by stiffness of the axial muscles with painful spasms. Affected people usually have high titers of anti-GAD and approximately one third to one half will develop type 1 diabetes [60].

Anti-insulin receptor antibodies can cause diabetes by binding to the insulin receptor thereby reducing the binding of insulin to target tissues [61]. However, these antibodies can also act as an insulin agonist after binding to the receptor and can thereby cause hypoglycemia [62]. Anti-insulin receptor antibodies are occasionally found in patients with systemic lupus erythematosus and other autoimmune diseases [63].

**Other genetic syndromes associated with diabetes**

Many genetic syndromes are accompanied by an increased incidence of diabetes mellitus. These include the chromosomal abnormalities of Down syndrome, Klinefelter syndrome, and Turner syndrome. These and other similar disorders are listed in Table 1.4.
Diabetes is commonly observed in cystic fibrosis patients. While it shares features of type 1 and type 2 diabetes, cystic fibrosis-related diabetes (CFRD) is a distinct clinical entity. It is primarily caused by insulin insufficiency, although fluctuating levels of insulin resistance related to acute and chronic illness and medications such as bronchodilators and glucocorticoids also play a role [64]. Since blood glucose levels within the IGT range appear to have an adverse effect on lung function, it has been suggested that diagnostic criteria for CFRD should be lower than that for other forms of diabetes, but data are currently inadequate to make this change [64]. CFRD is not associated with atherosclerotic vascular disease, despite the fact that individuals with cystic fibrosis nowadays can have a lifespan well into the 50s and 60s.

There are several distinct clinically defined subgroups of diabetes where an etiology has not yet been defined. In recognition of this, during the most recent WHO consultation, it was recommended that a category of “unclassified” or “nonclassical phenotype” be available.

### Diabetes in children and youth

Type 1 diabetes in children and youth is typically characterized by weight loss, polyuria, polydipsia, blurring of vision, very high plasma glucose concentrations, and ketonuria. The diagnosis is usually very clear with high random glucose values, and there is rarely a need to investigate with an oral glucose tolerance test (OGTT). Type 2 diabetes in children is associated with milder symptoms and is often associated with obesity. In these cases, diagnosis is made using any one of OGTT, fasting plasma glucose, or HbA1c, with preference for HbA1c as there is no requirement to fast. However, there is still debate as to the use of the latter in children [65].

Classification of diabetes in youth poses special problems. Although type 1 diabetes remains the most common form of diabetes in youth of European background, type 2 diabetes is increasingly common, especially among adults at particularly high risk of type 2 diabetes. With the increase in obesity over the last 20 years, there has been an increase in type 2 diabetes in children especially among ethnicities at high risk as well as an increase in the number of children with type 1 who are overweight. Type 2 diabetes may also be present in youth with ketosis or ketoacidosis, which serves only to compound the problem further. While a practical delineation between these may be the use of insulin, it can no longer be assumed that those on insulin are type 1. Other investigations which could provide insight include measurement of C-peptide, characteristic type 1 antibodies, for example anti-GAD antibodies, and the monitoring of endogenous insulin secretion over time [17].

There has also been an increase in the number of children and adolescents with a mixture of the two types of diabetes, that is, subjects who are obese and/or with signs of insulin resistance as well as being positive for markers of autoimmunity to β cells. These cases present a problem under the current classification as they present with an overlapping phenotype of both type 1 and type 2 diabetes and have been referred to as hybrid diabetes, double diabetes, or latent autoimmune diabetes in youth (LADY) [66]. In such children, presentation of double diabetes is similar to LADA in adults. However, unlike LADA, little is known about the prevalence of double diabetes or the prevalence and significance of autoimmune markers in children. In addition, whether autoimmune-positive youth with double diabetes progress more rapidly to insulin dependence than those with type 2 diabetes without is not known. This is particularly important as these children/youth could be at risk for complications associated with β-cell dysfunction, as well as macro- and microvascular complications of type 2 diabetes. It has been suggested that the current classification of diabetes should be revised to include this new phenotype [66].

Another challenge among young people is the possibility of misdiagnosis of monogenic diabetes as type 1 and type 2. As noted previously, monogenic diabetes results from the inheritance of mutation(s) in a single gene that regulates β-cell function or less commonly in genes related to insulin resistance.

The clinical characteristics of a child with monogenic diabetes compared to children and youth with type 1 and type 2 are shown in Table 1.5. Monogenic diabetes should be considered in a child initially diagnosed as type 1 who has been diagnosed at less than 6 months of age, has a family history of diabetes with a parent affected, evidence of endogenous insulin production outside the “honeymoon” phase of diabetes with detectable C-peptide, and the absence of pancreatic islet autoantibodies (measured at diagnosis) [67].

In children with an initial diagnoses of type 2, a diagnosis of monogenic diabetes should be considered in the following circumstances: when the child is not obese or other diabetic family members have weight in the normal range, and the child does not have acanthosis nigricans; when the child is from an ethnic group with a low prevalence of type 2 diabetes and when there is no evidence of insulin resistance with normal fasting C-peptide levels [24,68].

In scenarios when monogenic diabetes is misdiagnosed as type 1 or 2, the afore-mentioned criteria should be considered as a

---

**Table 1.4 Other genetic syndromes sometimes associated with diabetes**

<table>
<thead>
<tr>
<th>Syndrome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Down syndrome</td>
</tr>
<tr>
<td>Friedreich's ataxia</td>
</tr>
<tr>
<td>Huntington's chorea</td>
</tr>
<tr>
<td>Klinefelter syndrome</td>
</tr>
<tr>
<td>Lawrence–Moon–Biedl syndrome</td>
</tr>
<tr>
<td>Myotonic dystrophy</td>
</tr>
<tr>
<td>Porphyria</td>
</tr>
<tr>
<td>Prader–Willi syndrome</td>
</tr>
<tr>
<td>Turner syndrome</td>
</tr>
<tr>
<td>Wolfram syndrome</td>
</tr>
<tr>
<td>Others</td>
</tr>
</tbody>
</table>

---

### References


[68] Re 符号 missing in this reference.
Table 1.5 Clinical characteristics of type 1 diabetes, type 2 diabetes, and monogenic diabetes in children and adolescents [67]

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Type 1 Polygenic</th>
<th>Type 2 Polygenic</th>
<th>Monogenic Monogenic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age on onset</td>
<td>6 months and older</td>
<td>Usually pubertal (or later)</td>
<td>Often postpubertal except glucokinase and neonatal diabetes</td>
</tr>
<tr>
<td>Clinical presentation</td>
<td>Most often acute, rapid</td>
<td>Variable; from slow, mild (often insidious) to severe</td>
<td>Variable (may be incidental in glucokinase)</td>
</tr>
<tr>
<td>Autoimmunity</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Ketosis</td>
<td>Common</td>
<td>Uncommon</td>
<td>Common in neonatal diabetes, rare in other forms</td>
</tr>
<tr>
<td>Obesity</td>
<td>Population frequency</td>
<td>Increased frequency</td>
<td>Population frequency</td>
</tr>
<tr>
<td>Acanthosis nigricans</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Frequency (% of all diabetes in young people)</td>
<td>Usually &gt;90%</td>
<td>Most countries &lt;10%, Japan 60–80%</td>
<td>71–3%</td>
</tr>
<tr>
<td>Parents with diabetes</td>
<td>2–4%</td>
<td>80%</td>
<td>90%</td>
</tr>
</tbody>
</table>

whole rather than individually and are not absolute. DNA testing is now also available for diagnosis of monogenic diabetes.

**DIAGNOSTIC CRITERIA**

Diabetes is characterized by hyperglycemia, and thus diagnostic tests focus on establishing elevated blood glucose levels [69]. A casual blood glucose, fasting glucose or an OGTT of 75 grams may be performed. For children, the oral glucose load is proportional to body weight at 1.75 g per kg body weight. Recently, HbA1c has been added as an acceptable and reliable means of diagnosing diabetes (discussed later). The cutpoints for the diagnosis of diabetes are listed in Table 1.6.

In the absence of symptoms clearly attributable to diabetes, a diagnosis should not be based on a single measurement, but requires results within the diabetes range on two separate days.

The most notable change in diagnostic criteria in recent years is the recommendation by the ADA and WHO to use HbA1c for diagnosis of diabetes. A summary of the evolution of this decision is described in the following section.

**Diagnosis of diabetes using HbA1c**

HbA1c is a hemoglobin variant primarily composed of glycated hemoglobin, which is formed by the nonenzymatic attachment of glucose to hemoglobin [70]. It was first identified in 1968 by Kadar, who noted it was associated with diabetes. By 1980, its clinical utility as a marker of glycemic control had been recognized. By the 1990s, supported by strong evidence from two studies, the Diabetes Control and Complications Trial [71] and the United Kingdom Prospective Diabetes Study [72], and the development of new high throughput methods and improved coefficients of variation (CV), HbA1c had become the cornerstone marker in the monitoring of diabetes. In more recent years, a US national glycohemoglobin standardization program has been established and the International Federation of Clinical Chemistry (IFCC) has taken the lead to ensure that HbA1c assays are standardized. In 2011, the units of reporting were also changed from percentage points to IFCC mmol/mol. After a period of dual reporting, HbA1c will be reported in mmol/mol in many countries.

With some improvement in the assay and standardization of HbA1c, together with evidence from key trials demonstrating the importance of intensive glycemic control (as reflected by HbA1c levels) in reducing the risk of microvascular complications of diabetes, the move from using glucose for diagnosis to HbA1c had begun to gather support. However, concerns over standardization of the HbA1c assays and over other factors that may affect HbA1c continued to dampen the enthusiasm for use of HbA1c for diagnosis. In the last 10 years, however, several developments have resulted in the incorporation of HbA1c into the diagnostic armamentarium. There has been significant improvement in the assays of HbA1c [73], analysis from eight different studies showed that HbA1c is as strongly related to the presence of diabetic retinopathy as are blood glucose levels [74], and HbA1c is strongly predictive of macrovascular outcomes and mortality [75,76].

The advantages of using HbA1c for diagnosis are clear. Firstly, HbA1c has far less day-to-day biological variation than fasting or 2-hour glucose [77]. Secondly, HbA1c is stable for one week at room temperature after collection while glucose is susceptible to glycolysis despite the use of fluoride oxalate to preserve the sample. Thirdly, unlike glucose measurement, there is no requirement for the patient to fast. Finally, glucose
Table 1.6 Values for diagnosis* of diabetes and other categories of hyperglycemia using glucose [69] and HbA1c

<table>
<thead>
<tr>
<th>Glucose concentration (mmol L(^{-1}))</th>
<th>Plasma</th>
<th>Whole blood</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Venous</td>
<td>Capillary</td>
</tr>
<tr>
<td><strong>Diabetes mellitus:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fasting or 2-h post glucose load</td>
<td>≥7.0</td>
<td>≥7.0</td>
</tr>
<tr>
<td><strong>Impaired glucose tolerance:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fasting and 2-h post glucose load</td>
<td>&lt;7.0</td>
<td>&lt;7.0</td>
</tr>
<tr>
<td>Impaired fasting glucose:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fasting and (if measured) 2-h post glucose load</td>
<td>6.1–6.9</td>
<td>6.1–6.9</td>
</tr>
</tbody>
</table>

**Diagnostic cutpoints for pre-diabetes and diabetes using HbA1c**

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Diabetes</td>
<td>≥6.5%</td>
<td>(48 mmol mol(^{-1}))</td>
</tr>
<tr>
<td>Increased risk (ADA) [5]</td>
<td>5.7–6.5%</td>
<td>(38–48 mmol mol(^{-1}))</td>
</tr>
<tr>
<td>High risk (International Expert Committee) [79]</td>
<td>6.0–6.5%</td>
<td>(42–48 mmol mol(^{-1}))</td>
</tr>
</tbody>
</table>

*Note that diabetes can only be diagnosed in an asymptomatic individual when these diagnostic values are confirmed on another day. Ranges of values are inclusive, i.e., 6.1–6.9 means ≥6.1 and <7.0.

levels are also susceptible to modification by short-term lifestyle intervention while HbA1c reflects glycemia over a period of 3 to 4 months.

The major disadvantage of HbA1c is that there are a number of nonglycemic conditions that interfere with the assay. In particular, alterations of red blood cell turnover (e.g. kidney failure, hematocrit deficiencies, hemolysis, acute blood loss, pregnancy, and erythropoietin therapy) may affect the relationship between HbA1c and recent glycemia. The other important disadvantage is the need for a laboratory to use an IFCC aligned assay and be part of a standardization program, which may not be possible in developing countries.

Cutpoints of HbA1c have been set using similar methods to those adopted for the setting of blood glucose criteria. Cross-sectional data from 47,364 individuals from 12 countries reported that the threshold for diabetes-specific retinopathy was 6.3% (45 mmol mol\(^{-1}\)), with an optimal decision limit of 6.5% [74]. This latter cutpoint has been adopted by the ADA and WHO as an appropriate cutpoint for diabetes.

Although support for the use of HbA1c for diagnosis of diabetes has increased over the years, several questions about its suitability remain. For example, what should be the appropriate HbA1c ranges for pre-diabetes or intermediate glycemia? The ADA suggested that 5.7–6.5% (39–48 mmol mol\(^{-1}\)) should be used [5] to indicate intermediate glycemia while the WHO [78] suggested that levels of HbA1c below 6.5% may indicate intermediate glycemia but were reluctant to indicate a precise lower cutpoint. An international expert committee suggested that those with HbA1c between 6.0–6.5% (42–48 mmol mol\(^{-1}\)) could be considered at high risk, and should be targeted for diabetes prevention activities [79].

A further concern about moving from glucose to HbA1c to diagnose diabetes is that we will observe a change in prevalence of diabetes, as an elevated HbA1c does not identify exactly the same individuals as does an elevated blood glucose. It should, however, be noted that a similar discrepancy in individuals identified also applies to diagnosis by fasting glucose compared to diagnosis by 2-hour plasma glucose in the OGGT.

In general, the use of HbA1c for diagnosis of diabetes results in a lower prevalence of diabetes with the magnitude of the difference between blood glucose-based prevalence and HbA1c-based prevalence varying widely between populations [80].

Diagnosis of diabetes using HbA1c is now recommended by both the ADA and WHO as detailed in Table 1.6. As discussed earlier, it is important to ensure that the HbA1c assay used meets stringent quality assurance test and is aligned with the IFCC standardization program. It is also important to ensure that there are no clinical conditions that preclude its accurate measurement.
Diagnosis of gestational diabetes mellitus

The diagnosis of gestational diabetes mellitus has been traditionally based on glucose tolerance levels measured between 24 and 28 weeks of gestation [7] using an OGTT. These guidelines have been modified due to the evidence from the Hyperglycemia and Adverse Pregnancy Outcome Study (HAPO) [81]. HAPO was a large, prospective, blinded, multinational study showing a strong and continuous relationship of maternal glycemia at 24–28 weeks with neonatal outcomes of increased birth weight and increased cord–blood serum C-peptide levels, and to increased cesarean section delivery rates in the mothers. Based on these data, new GDM guidelines [82] were proposed which have since been adopted internationally. Diagnostic criteria for gestational diabetes are shown in Table 1.7.

OTHER GLUCOSE TOLERANCE CATEGORIES

Impaired glucose regulation (impaired glucose tolerance and impaired fasting glycemia)

Impaired glucose tolerance (IGT) and impaired fasting glycemia (IFG) are categorized as stages in the natural history of disordered carbohydrate metabolism. They occur in all individuals as they progress from normal to diabetes, but since the transition through these states is rapid in type 1 diabetes, they are rarely identified in such individuals. Therefore, nearly all of the literature dealing with IGT and IFG is concerned with issues relating to type 2 diabetes, such as risk of developing type 2 diabetes and CVD.

IFG and IGT represent a metabolic state intermediate between normal glucose homeostasis and diabetes. The pathophysiologic aspects of hyperglycemia of each category are somewhat different. IGT is associated with muscle and liver insulin resistance and thus IGT is often associated with the metabolic or insulin resistance syndrome [7], while IFG is usually related to insulin secretory deficits.

A meta-analysis suggested that there is a positive relationship between IFG/IGT and diabetes which varies across ethnicity and age [83]. This study showed that individuals with combined IGT and IFG had the highest risk of future diabetes. In terms of sensitivity and specificity for the subsequent development of diabetes, the sensitivity of IFG as originally defined at 6.1 mmol L$^{-1}$ (110 mg dL$^{-1}$) is less than that of IGT in most populations [84], but the specificity of IFG is greater [85]. IGT is more common than IFG using 6.1 mmol L$^{-1}$ (110 mg dL$^{-1}$) in most populations but it should be noted that the sensitivity and specificity of both IGT and IFG is entirely dependent on the cutpoints selected, and not on any inherent differences between FPG and 2hPG [86]. If IGT and IFG are defined such that they have similar prevalence to each other, they then have the same predictive values for subsequent diabetes [86]. Further, with the ADA cutpoint of IFG (5.6 mmol L$^{-1}$ or 100 mg dL$^{-1}$ — see later), the sensitivity of IFG is similar to IGT, but the specificity falls.

Thus, neither the risk of developing diabetes nor the sensitivity and specificity for future diabetes seem to differ enough between IGT and IFG to suggest one category is more useful than the other. In reality, in most populations IGT is more prevalent than IFG (if IFG is defined as FPG of 6.1–6.9 mmol L$^{-1}$ (110–125 mg dL$^{-1}$)), and thus it identifies a greater proportion of those who will develop diabetes. Furthermore, although at the lower cutpoint of IFG of FPG 5.6–6.9 mmol L$^{-1}$ (100–125 mg dL$^{-1}$), the prevalence of IFG approaches that of IGT, the two groups remain limited in their overlap. Relying on only a FPG will not identify the same proportion of individuals at risk compared to undertaking an OGTT.

The relationship between IGT and IFG and CVD is well studied in meta-analyses. In a review of the evidence from 27 studies [87], IFG (at both cutpoints) and IGT were both associated with a significantly increased risk of approximately 20% of CVD.

Diagnosis of IGT and IFG categories has been traditionally made by measuring blood glucose levels, either in the fasting state (for IFG) or during an OGTT (for IGT) (see Table 1.6 for cutpoints). Since individuals with IFG may have diabetes, it is recommended that those who are found to have IFG should have an OGTT to exclude diabetes [7].

Whilst IGT has been part of the classification of glucose intolerance for many years, IFG was only added in 1997, with a lower cutpoint of 6.1 mmol L$^{-1}$ (110 mg dL$^{-1}$). However, in 2002, the ADA proposed a new cutpoint of IFG of 5.6 mmol L$^{-1}$ (100 mg dL$^{-1}$), as this maximized the sensitivity and specificity for predicting future diabetes [88]. On review of the same evidence, the WHO decided not to adopt this new cutpoint, as it significantly increased the number of people being labeled as abnormal, but without evidence that so doing would improve outcomes [69].

The purpose of defining other categories of glucose intolerance or prediabetes is to identify a group of the population at increased risk for the development of both diabetes and CVD, so that interventions (lifestyle and pharmacologic) can be applied to reduce these risks. IGT and IFG are considered risk factors for diabetes and CVD.

In summary, longitudinal data show that IFG and IGT are rather similar to each other in their ability to predict future diabetes and CVD. However, since the populations of

---

Table 1.7 Diagnostic criteria for gestational diabetes mellitus [82]

<table>
<thead>
<tr>
<th>Test Type</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fasting venous PG</td>
<td>≥5.1 mmol L$^{-1}$</td>
</tr>
<tr>
<td>1-h venous PG</td>
<td>≥10.0 mmol L$^{-1}$</td>
</tr>
<tr>
<td>2-h venous PG</td>
<td>≥8.5 mmol L$^{-1}$</td>
</tr>
</tbody>
</table>

One or more of these values must be abnormal for the diagnosis of GDM.
IFG and IGT have limited overlap with each other, undertaking the OGTT to identify those with IGT provides the opportunity to identify a greater proportion of the at-risk population.

**Normoglycemia**

The notion underpinning setting a normal category of glucose is that people with values below the upper limit of normal are at no or only "normal" risk of developing diabetes or its micro- and macrovascular complications [5,7,84,89]. Since the risks of future development of diabetes and CVD are related to blood glucose across most of its spectrum, and well into any normal ranges that have been set, such notions of "normal" blood glucose should be interpreted very cautiously. The actual setting of the cutpoint indicating normoglycemia over the years has undergone considerable changes. The early classification in 1985 by US NDDG that was adopted by WHO had set diabetes at a fasting glucose of 7.8 mmol L\(^{-1}\) (140 mg dL\(^{-1}\)) and those under this threshold were labeled "normoglycemia". In 1997, upon the availability of new data, the ADA, with support from WHO reset the cutpoint of a "normal" fasting plasma glucose from 7.8 mmol L\(^{-1}\) (140 mg dL\(^{-1}\)) to 6.0 mmol L\(^{-1}\) (110 mg dL\(^{-1}\)) [5,7,89]. In 2002–2003, the ADA recommended that the cutpoint be 5.5 mmol L\(^{-1}\) (100 mg dL\(^{-1}\)) [88].

**Summary**

The classification of diabetes is an evolving process. As the research into diabetes is a continuing and dynamic process and epidemiologic and clinical studies are in progress, there may well be revision and refinement of the classification system. As more knowledge emerges about the etiology of cases currently positioned in the type 2 process category, modification and refinement may be necessary.

**References**


CHAPTER 2

Epidemiology and risk factors for type 1 diabetes mellitus

Andrea K. Steck and Marian J. Rewers
Barbara Davis Center for Childhood Diabetes, University of Colorado Denver, Aurora, CO, USA

Key points

• The incidence of type 1 diabetes has been increasing progressively over the last half century by 3–5% per year.
• Type 1 diabetes results from a chronic autoimmune destruction of the pancreatic β cells with a preclinical period marked by the presence of autoantibodies to pancreatic β-cell antigens.
• Though genetic markers can identify varying risk, it is only once autoimmunity has begun that a high positive predictive value is achieved.
• Multiple islet autoantibodies are present in the great majority of prediabetics.
• Studies indicate that genetic factors alone cannot explain the etiology of type 1 diabetes.
• Seasonality, increasing incidence and epidemics of type 1 diabetes suggest a critical role of environmental factors, such as infections with certain viruses and effects of early childhood diet.
• In the absence of renal disease, the long-term mortality risk in type 1 diabetes is not increased compared with the general population.

Introduction

Type 1 diabetes (T1DM) is one of the most prevalent severe chronic diseases of childhood, affecting more than 170,000 children in the United States, an increase of 23% since 2001 [1]. In the US, more than 25,000 children are diagnosed annually with 1:200 children and 1:100 adults diagnosed with T1DM during the lifespan [2]. T1DM is the leading cause of end-stage renal disease, blindness, and amputation, and a major cause of cardiovascular disease and premature death in the general population. Annually, in the US, an estimated 70–200 children die at the onset of diabetes [3] with 30% of children who develop diabetes presenting with ketoacidosis. For unknown reasons, the incidence of T1DM has been increasing progressively over the last half century by 3–5% per year [4]. In addition the percentage of children expressing the highest risk HLA genotype (DR3/4-DQ2/8) has dramatically decreased over the past 50 years [5], apparently reflecting increased disease penetrance of lower risk haplotypes. The high incidence, associated severe morbidity, mortality, and associated healthcare expenditures make T1DM a prime target for prevention. Population-based epidemiological studies as well as family studies and clinical trials have provided new insights into the pathogenesis and natural history of T1DM. Such studies are essential for appropriate diagnosis and for evidence-based programs of prevention and treatment.

Prevalence and incidence

The prevalence of T1DM, that is the proportion of people in the population who have the disease at a given point in time, is determined not only by disease incidence but also by case survival, which may vary markedly in populations. The prevalence of β-cell autoimmunity appears to be roughly proportional to the incidence of T1DM in different populations. In contrast, the prevalence of β-cell autoimmunity in first-degree relatives of T1DM persons does not differ a lot between high and low risk countries for T1DM. The prevalence of T1DM in children aged less than 15 years ranges from 0.05 to 0.3% in most European and North American populations [6].

Incidence is the rate at which new cases of disease appear in the population and is usually expressed as the annual number of new cases per 100,000 persons. Incidence of T1DM varies by geographic location, ethnicity, age, gender, and time. The incidence of T1DM is increasing worldwide both in low and high incidence populations (Figure 2.1) [7]. The average annual incidence is the...
highest in Finland where it has increased to 64 per 100,000 per year in 2005 [8] while the current mean incidence is 18/100,000 person-years in Australia [9].

**Geographic location**

One of the most striking characteristics of T1DM is the large geographic variability in the incidence [10–13]. The incidence rates of T1DM vary from 0.73/100,000 per year in China [14] to 60/100,000 per year in Finland. A child in Finland is 100 times more likely to develop diabetes than a child in China. A clear difference appears between the Northern and Southern Hemisphere with countries below the equator having a lower incidence; in contrast above the equator the disease is common. The largest intracontinental variation in incidence appears to be in Europe. There are also noticeable within-country variations in incidence rates, which can only be partially explained by racial composition of the population.

The geographic and ethnic variation in T1DM risk may reflect either different pools of susceptibility genes or different prevalence of causative environmental factors or a combination of both.

**Race and ethnicity**

There are also striking racial differences in T1DM risk within the same population, although they are not as important as the geographic differences. In the United States, non-Hispanic whites are about one and a half times as likely to develop T1DM as African Americans or Hispanics [15,16]. There are similar differences reported in Europe, where France has lower rates than do Britain and Scandinavia.

Interestingly, there is also evidence that migrants from a country with low incidence of T1DM soon have the higher rate incidence of the country where they live; for example South Asian children living in the UK have similar overall rates of T1DM compared to indigenous white children [17]. Genetic factors are unlikely to explain such a rapid change, implying an influence of environmental factors in disease etiology.

**Age and gender**

T1DM incidence peaks at the ages of 2, 4–6 and 10–14 years, perhaps due to alterations in the pattern of infections or increases in insulin resistance. Few cases of T1DM develop in the first year of life. The main incidence peak occurs at puberty, with females having a pubertal peak about 1 year earlier. There is a decrease in incidence after puberty for both sexes [18]. In Sweden, the age-specific incidence rates vary from 5–20/100,000 person-years in adults aged 15–35 years and decline with age. Several studies indicate a stable or decreasing trend of T1DM among young adults [18], while, in contrast, a study from Italy (lower-risk population) found that the incidence from 1984 to 1996 increased not only in children but also in young adults [19].

In general, males and females have similar risk of T1DM [20]. In lower-risk populations, such as Japan or African Americans, there is a female preponderance, while in high-risk groups, there is a slight male excess [18]. Interestingly, even within Europe, populations with an incidence higher than 20/100,000 (Sardinia, UK, Italy, Finland, Norway) had male excess, whereas those with a lower rate (the Baltic countries, Macedonia, Yugoslavia, Romania) had female excess [21].

**Increasing incidence and seasonality**

There is evidence for marked variations in the incidence of T1DM over time, both seasonally and annually. In the Northern Hemisphere, the incidence declines during the warm summer months; similarly in the Southern Hemisphere, the seasonal pattern exhibits a decline during the warm months of December and January, implicating a climatic factor. This seasonal pattern appears to occur predominantly in older children [16], suggesting that factors triggering diabetes may be related to school attendance. Most population-based registries have shown increasing T1DM incidence over time [22,23]. Several studies have observed periodic outbreaks superimposed on a steady secular increase in incidence.

While the increase in T1DM incidence has affected all age groups, several studies have reported a particular increase among the youngest children [24]. In the youngest age group (0–4 years) an increase of 11% and 24% per year was reported in, respectively, the UK and Switzerland [25]. On the contrary, some studies show the highest increase in incidence in children 10–14 years old [13].

In the EURODIAB ACE Study Group, representing most European countries and Israel, the rates of increase were 6.3% (4.1–8.5%) for children aged 0–4 years, 3.1% (1.5–4.8%) for 5–9 years, and 2.4% (1.0–3.8%) for 10–14 years [23].

**β-Cell autoimmunity and risk factors**

Type 1A diabetes results from a chronic autoimmune destruction of the pancreatic β cells, probably initiated by exposure of a genetically susceptible individual to some environmental
agent(s). This preclinical period is marked by the presence of autoantibodies to pancreatic β-cell antigens such as insulin, GAD65 (Glutamic Acid Decarboxylase), ICA512 (called also IA-2) or ZnT8 (Zinc Transporter 8), and precedes the onset of hyperglycemia by a few years. Several prospective studies have reported that these autoantibodies can appear early in childhood and the presence of two or more of these antibodies is highly predictive for the development of diabetes [26]. However, the etiology of the autoimmune process and β-cell destruction is not known. The islet cell antibody (ICA) assay, using immunofluorescence and pancreatic tissue has been notoriously difficult to standardize and has been replaced by a combination of radioassays for autoantibodies to insulin [27], GAD and IA-2. These tests have been shown to be quite sensitive and predictive in relatives of T1DM patients [26] and in the general population. Several prospective cohort studies investigate early genetic and environmental factors contributing to risk of β-cell autoimmunity, and progression to diabetes in relatives of patients with T1DM and in a general newborn population [28,29].

Prediction of T1DM during the preclinical period offers opportunities for prevention of T1DM. Prediction at present is based upon genetic, immunologic, and metabolic information (Figure 2.2). Though genetic markers can identify varying risk, it is only once autoimmunity has begun (marked by the presence of multiple autoantibodies to pancreatic β-cell antigens such as insulin, GAD65, IA-2 or ZnT8) that a high positive predictive value (>90%) can be achieved, and multiple autoantibodies are present in the great majority of prediabetics [30,31].

Family history of T1DM
Concordance rates for T1DM in monozygous twins (MZ) with long-term follow-up is greater than 50% [32], compared to 6–10% in dizygous (DZ) twins, which is similar to what is found in non-twin siblings. With long-term (>30 years) follow-up, at least 2/3 of initially discordant MZ twins develop persistent β-cell autoantibodies and/or diabetes.

Among first-degree relatives, siblings are at a higher risk (5–10% risk by age 20) than offspring; offspring of diabetic fathers are at a higher risk (~12%) than offspring of diabetic mothers (~6%) [33]. It is not clear why offspring of diabetic mothers are at lower risk compared to offspring of fathers, with one study reporting that presence of transplacental islet autoantibodies may decrease risk [34]. It is very likely that both environmental factors and genetic susceptibility are essential for development of T1DM, but the environmental factors may be ubiquitous and thus not be a major determinant of which individual develops diabetes given genetic susceptibility.

### HLA class II susceptibility genotypes
The estimated risk of developing T1DM for general population children is 1/300 while the risk for children who have the HLA-DR3/4,DQB1*0201,DQB1*0302 genotype is approximately 1:15–1:25 (Table 2.1) [35]. Only 2.4% of the general population carries this genotype compared to 25–40% of T1DM patients. Deschamps and coworkers examined the predictive value of HLA typing in a study of 536 siblings of diabetic probands in France [36]. The risk of T1DM after 8 years, estimated by life table analysis, was 10% for siblings who were HLA identical with the probands, 3–4% for siblings with either DR3 or DR4, and 16% for those with HLA-DR3/DR4 genotype. Recently, there has been evidence for additional susceptibility loci within or linked to the MHC, independent of HLA-DR/DQ, such as HLA class I alleles [37,38]. In the DAISY (Diabetes Autoimmunity Study in the Young) cohort, siblings of children who have HLA-DR3/DR4-DQB1*0302 and are identical by descent for both HLA haplotypes with their diabetic proband sibling had a 65% risk for developing islet autoantibodies by age 7 years and a 50% risk of developing diabetes by age 10 years [39]. This strongly suggests that additional (non-DR,DQ) MHC-linked genes determine T1DM risk. On the other hand, DAISY general population children with DR3/DR4-DQB1*0302 who lack protective alleles DPB1*0402 and DRB1*0403 have a risk of 20% of activating islet autoimmunity [40]. Only 25% of autoantibody positive new onset children presenting to the Barbara Davis Center for Childhood Diabetes in Denver during this past decade have the HLA-DR3/4 genotype [41].

### Table 2.1 Risk, by the age of 20 years, of type 1 diabetes and β-cell autoimmunity in the general population and family members of T1DM patients

<table>
<thead>
<tr>
<th>Risk Group</th>
<th>Type 1 Diabetes</th>
<th>Pre-Diabetic Autoimmunity</th>
</tr>
</thead>
<tbody>
<tr>
<td>General population</td>
<td>All HLA genotypes</td>
<td>1:300</td>
</tr>
<tr>
<td></td>
<td>HLA-DR3/4,DQB1*0302</td>
<td>1:15</td>
</tr>
<tr>
<td>Family members</td>
<td>Maternal offspring</td>
<td>1:50</td>
</tr>
<tr>
<td></td>
<td>Paternal offspring</td>
<td>1:15</td>
</tr>
<tr>
<td></td>
<td>Siblings (all)</td>
<td>1:12–1:35</td>
</tr>
<tr>
<td></td>
<td>Monozygotic twins</td>
<td>1:3</td>
</tr>
<tr>
<td></td>
<td>HLA-identical siblings</td>
<td>1:4</td>
</tr>
</tbody>
</table>

Several HLA haplotypes dominantly protect, including HLA-DQA1*0102,DQB1*0602, DRB1*1401, DQA1*0101, DQB1*0503 and DRB1*0701,DQA1*0201,DQB1*0303 [42]. Too few HLA-DQB1*0602 individuals expressing multiple islet autoantibodies have been studied to know the exact magnitude of protection once autoantibodies are present, but 1% of new-onset patients expressing islet autoantibodies have DQB1*0602 versus 20% of the general US population.

Other genetic factors associated with type 1 diabetes

More than 40 non-HLA susceptibility gene markers have been confirmed [43]. At present, polymorphisms of the insulin gene and PTPN22 gene contribute most to diabetes risk after HLA alleles [44]. Adding high-risk alleles of these genetic markers to HLA Class II genotyping can increase risk, but even for these loci with odds ratios between 1.7 and 2.0, the effect is small.

Recent studies indicate that specific single nucleotide polymorphisms (e.g. PTPN22) [45] of non-MHC genes are associated with dramatic differences in T-cell signaling. For instance, the PTPN22 allele associated with diabetes risk is associated with decreased T cell-receptor signaling (gain of function). It is known that a subset of individuals without the specific polymorphism have similar phenotypes, and it is possible that, as diabetes-associated lymphocyte phenotypes are defined, they will contribute to disease prediction [46].

Islet autoantibodies

The first large-scale studies of the prediction of T1DM relied upon the detection of cytoplasmic islet cell autoantibodies (ICA) assays based on indirect immunofluorescence. High titer cytoplasmic ICA is most often associated with the presence of multiple islet autoantibodies and therefore a high risk of progression to diabetes [47]. Screening for risk of T1DM now utilizes “biochemical” autoantibody assays for specific islet autoantigens. These include autoantibodies to insulin (IAA), glutamic acid decarboxylase (GAD), IA-2 (ICA512) and most recently ZnT8 [48]. Individuals having a single positive autoantibody (insulin, GAD, IA-2, or ZnT8 autoantibodies) are at low risk for progression to T1DM. Single positive autoantibodies can be a nonreproducible false positive result (e.g. switched sample), transient, or represent the presence of an autoantibody reacting with the specific autoantigen that does not confer increased risk of diabetes (e.g. low affinity insulin autoantibodies) [49]. Individuals expressing two or more positive autoantibodies, especially on multiple tests over time, are at very high risk of progressing to diabetes. This high risk may result from autoimmunity spreading to other autoantigens whose targeting increases destruction or from the high statistical specificity of expression of multiple autoantibodies. The Diabetes Autoantibody Standardization Program (DASP) workshop aims to improve and standardize measurement of autoantibodies associated with T1DM across laboratories [50].

Different islet autoantibodies have been associated with different risks of progression, with IA-2 autoantibodies most often associated with expression of other biochemical autoantibodies and high risk [51]. Of note, insulin autoantibodies are extremely high at the onset of diabetes in young children while usually negative in individuals first presenting with diabetes after age 12. There is a log-linear inverse relationship between the levels of insulin autoantibodies and the age of onset of diabetes [52]. In DAISY, 89% of children who progressed to diabetes expressed ≥2 autoantibodies with cumulative incidence of 74% by age 10 for those expressing three autoantibodies (Figure 2.3). In children expressing ≥2 autoantibodies, there is no significant difference in progression to diabetes between relatives and general population subjects.

Autoantibody screening among relatives

The initial large screening studies found approximately 3% of first-degree relatives to be positive for ICA. With analysis of biochemical autoantibodies, it has become evident that the presence of ICA in the absence of GAD65 or ICA512 autoantibodies is associated with a low risk of progression to diabetes [26,53]. The cumulative risk of developing diabetes within 15 years is only 2.8% for individuals with ICA but without GAD or ICA512 (IA-2) autoantibodies versus 66% for those with ICA and either or both GAD or ICA512 autoantibodies.

As in many studies, in the DAISY study, the incidence of islet autoantibodies is much higher in first-degree relatives, compared to the general population. The risk by age 10 is particularly high in the HLA-DR3/4,DQB1*0302 positive siblings (43%) and offspring (34%) and moderate-risk siblings (19%) [54]. In the DAISY children without a first-degree relative with T1DM, the incidence of persistent islet autoantibodies by the age of 9 years is 10.6%, 5.5% and 3.4% in, respectively, high-, moderate- and average-HLA risk groups. The HLA genotype and having a diabetic relative, but not gender or ethnicity, predict development of islet autoantibodies.
Islet autoantibodies appear early in life. The BABYDIAB, DIPP, and DAISY studies have demonstrated that a significant proportion of first-degree relatives progressing to T1DM before age 15 develop islet autoantibodies before their 2nd birthday [31,55,56]. Although IAA usually appear first, a significant percentage of children followed from birth initially express GAD65 autoantibodies, while IA-2 and ZnT8 autoantibodies usually develop later.

**Autoantibody screening in the general population**

Islet autoantibody screening studies have included mainly first-degree relatives; however, 90% of T1DM cases occur in individuals with no family history of T1DM. Several ongoing studies have followed children from birth for the development of islet autoantibodies. The two studies with the longest follow-up are the DAISY study from Denver Colorado and the DIPP study in Finland [57]. More recently, the multicenter international TEDDY study (The Environmental Determinants of Diabetes in the Young) [58] has enrolled over 8600 high-risk general population children identified by newborn screening into a prospective follow-up for islet autoantibody measurements from birth.

While the risk of developing islet autoantibodies is 3–4 times higher in relatives than in the general population with the same HLA-DR/DQ genotypes, the persistent presence of islet autoantibodies portends similar high risk of progression to diabetes for both relatives and the general population [52,59]. Higher titer and higher affinity of the autoantibodies as well as presence of autoantibodies to multiple autoantigens predict higher risk of diabetes [60].

**Measurement of autoantibodies in adult-onset diabetes**

After the age of about 15 years, measuring the incidence of T1DM is complicated by misclassification of some patients as type 2 diabetics. It has been estimated that at least 37% of T1DM is diagnosed after the age of 19 years and 15% after the age of 30 years [61]. New cases of T1DM presenting in adult life tend to have a longer duration of symptoms before diagnosis and higher C-peptide levels remaining at diagnosis compared with those presenting in childhood, suggesting a slower rate of β-cell destruction. However, C-peptide levels in islet autoantibody positive adult diabetic patients are still significantly lower than in those with type 2 diabetes. Assays for GAD65 autoantibodies have been the most useful in identifying latent autoimmune diabetes in adults (LADA).

Between 5 and 10% of patients with a diagnosis of gestational diabetes have positive islet autoantibodies and the great majority progress to T1DM [62]. Among adults diagnosed with type 2 diabetes and participating in the UKPDS, the proportion of patients with ICA and GAD65 decreased with increasing age at diagnosis: from 21% in patients aged 25–34 to 4% in those aged 55–65 for ICA and from 34% to 7% for GAD65 [63]. Most (94%) of patients with ICA and 84% of those with GAD65 required insulin therapy by 6 years, compared with 14% of those without the antibodies.

**Measurement of β-cell function**

Direct measurement of the functional mass of islet cells is currently not possible. The first-phase insulin response (FPIR) to intravenous glucose can help predict progression to diabetes among individuals with positive islet autoantibodies [64,65]. The FPIR is usually calculated as the sum of the 1- and 3-min insulin levels measured after glucose is administered intravenously at 0.5 g per kg body weight. Low FPIR has been defined as below the 1st, 5th or 10th percentile of the distribution in nonobese healthy subjects. Subjects from the Joslin family study with FPIR below the first percentile on the first test had a 3-year diabetes-free survival rate of 13% compared with 78% for the group with higher FPIR [66]. The FPIR is severely depressed at the time of detection of islet autoantibodies in many young children. In the DPT-1 study, loss of the FPIR was strongly associated with diabetes [64]; the average time from the fall of FPIR below the 1st percentile to the onset of diabetes was 1.8 years. A recent study from the Belgian registry reported similar results relative to first-phase insulin secretion with hyperglycemic clamps amongst individuals with IA-2 autoantibodies [67].

The oral glucose tolerance test (OGTT), performed in clinical trials of T1DM prevention largely to formalize the diagnosis of clinical diabetes, has long been known to have some value in predicting progression to T1DM among subjects with islet autoantibodies [68]. DPT-1 reported that a risk score based on age, BMI, and OGTT indexes (glucose and C-peptide values), without use of IVGTTs or additional autoantibodies, accurately predicted T1DM in ICA-positive relatives [69]. The likelihood of progression to diabetes increased with mild fasting or post oral glucose-load dysglycemia.

In 1993, Leech and coworkers reported that HbA1c measured by high-performance liquid chromatography may be slightly but significantly higher in non diabetic ICA-positive teenagers compared to ICA negative controls [70]. More recently, the DAISY study has demonstrated that Hb1Ac steadily increases within the normal range over a few years preceding diabetes onset and may therefore be useful in early detection of T1DM [71]. Among children who have persistent islet autoimmunity, increase in HbA1c predicted increased risk of progression to T1DM, with a hazard ratio of 4.8 for each 0.4% (SD) increase in HbA1c, independent of random glucose and number of autoantibodies.

Recently, increased HbA1c level has been added by the joint American Diabetes Association (ADA), International Diabetes Federation, and European Association for the Study of Diabetes International Expert Committee (IEC) as a diagnostic tool for diabetes with a recommended threshold of 6.5% for diagnosis of diabetes on two tests [72]. This threshold was established based on research conducted in adults with type 2 diabetes. However, studies in adolescents at high risk for T1DM suggest that HbA1c > 6.5% is a specific but not sensitive early indicator for T1DM [73].
Through the Diabetes Complications and Control Trial (DCCT), it has been shown that patients who have sustained production of C-peptide have lower rates of severe hypoglycemia, microalbuminuria, and retinopathy [74,75]. In the Diabetes Prevention Trial-Type 1 (DPT-1), post challenge C-peptide levels begin to decrease appreciably in the 6 months before diagnosis and continue to decrease within 3 months after diagnosis [76]. Recent data from the TrialNet study show a biphasic decline in C-peptide during the first 2 years post diagnosis, with a higher rate of fall during the first year [77].

Environmental factors
Twin [78] and family studies indicate that genetic factors alone cannot explain the etiology of T1DM. Seasonality, increasing incidence and epidemics of T1DM as well as numerous ecological, cross-sectional and retrospective studies suggest a critical role of environmental factors, such as infections with certain viruses (especially enteric infections in early life) and effects of early childhood diet. Natural history studies that follow children at increased risk of T1DM provide the best opportunity to study environmental triggers.

Viruses: Herpes viruses, mumps, rubella, and retroviruses [79] have been implicated. Viral infections appear to initiate autoimmunity and perhaps also precipitate diabetes in subjects with autoimmunity. ICA or IAA have been detected after mumps, rubella, measles, chickenpox, Coxsackie, ECHO-4, and rotavirus [80] infections.

While several viruses have been linked to T1DM, studies have provided the strongest overall evidence for entroviruses, although results have been somewhat conflicting [81]. In one longitudinal birth cohort study (DAISY), progression from islet autoimmunity to T1DM seemed to increase after an enterovirus infection [82]. In the Finnish type 1 Diabetes Prediction and Prevention (DIPP) study, enteroviral infections are seen more frequently in prediabetic children and prior to the onset of islet autoimmunity, implying a temporal relationship between enterovirus infections and the induction of β-cell autoimmunity [83]. Although several studies report that enteroviruses may play a role in the pathogenesis of T1DM, the evidence that enterovirus infections are associated with initiation or progression of islet autoimmunity is still inconsistent [84].

Effect of childhood infections and daycare exposure: Early infectious exposure may play a role in the development of immunoregulatory mechanisms that protect against diabetes. Social mixing through attendance at daycare in early infancy appears to confer protection against the development of childhood diabetes [85]. Although several other well-designed case-control studies show a statistically significant protective effect of day care exposure on T1DM [86], meta-analysis reveals too much heterogeneity to accept overall synthesis [87].

Improvement in hygiene: Genetic models are unable to explain the apparent temporal changes in the incidence of T1DM [88]. Alternative explanations look at environmental factors and some invoke the congenital rubella model. Briefly, increased hygiene in the Western world has led to a decline in immunity to common infections among women in child-bearing age. These women are more likely to develop viremia during pregnancy resulting in congenital persistent infection of β cells and early onset T1DM in the offspring. This model could explain both the increasing incidence of diabetes and the decreasing age of disease onset.

Routine childhood immunization: Neither type nor quantity or timing of vaccinations, including BCG vaccine, HIB vaccine, diphtheria, tetanus and pertussis vaccine, measles, mumps and rubella vaccine, hepatitis B vaccine, varicella vaccine, or tick-born encephalitis vaccine, have been associated with the development of islet antibodies and diabetes [86,89,90]. At least two studies even showed a possible protective effect of the measles-mumps-rubella vaccine [91].

Perinatal factors: Environmental risk factors may play a role early in life, possibly in utero. Several studies have investigated perinatal determinants for developing T1DM. Offsprings of a T1DM parent have an increased risk of developing diabetes, the risk being higher if the father has diabetes. Although the relation of maternal age and birth order to risk of T1DM is complex, several studies found that maternal age over 35–40 years [92] and/or increasing birth order [93] were associated with an increase in T1DM. There also seems to be a relatively weak but significant association between increasing birth weight and increasing risk of T1DM [94,95], although several other case-control studies have not found any association [96]. Interestingly, several studies suggest that early weight gain and/or rapid linear growth are risk factors for development of not only type 2 but also T1DM in children [97].

Dietary factors: Cow’s milk or wheat introduced at weaning trigger insulitis and diabetes in animal models perhaps through a molecular mimicry [98]. Breast-feeding may be viewed as a surrogate for the delay in the introduction of diabetogenic substances present in formula or early childhood diet. Several human studies suggested an association between short duration of breast-feeding and increase in T1DM [99,100], while cohort studies failed to find an association between cow’s milk and β-cell autoimmunity [90,101] (Figure 2.4). Interestingly, a study from Finland suggested that current cow’s milk consumption was more closely linked to prediabetic autoimmunity and diabetes than infant exposure [102]. To resolve this controversy, a dietary intervention trial to prevent T1DM by a short-term elimination of cow’s milk from infant diet (TRIGR) is underway [103].

Concerning the introduction of cereals, the DAISY [104] and the German BabyDiab [105] studies found that early introduction of gluten before the age of 3 months increases the risk of development of β-cell autoimmunity (Figure 2.4). However, delaying gluten exposure until the age of 12 months in the BabyDiet cohort did not substantially reduce the risk for islet autoimmunity in genetically at-risk children [106].
Vitamins and dietary supplements: Studies in vitro have shown that vitamin D3 is immunosuppressive or immunomodulating and studies in experimental models of autoimmunity have shown vitamin D to be protective [107]. Results from clinical studies have been somewhat conflicting. While a European study reported a protective effect of vitamin D supplementation in infancy against T1DM [108, 109], the DAISY cohort study did not show any association between vitamin D intake or 25(OH)D levels throughout childhood with the risk of islet autoimmunity or progression to T1DM [110] (Figure 2.4). On the other hand, maternal intake of vitamin D through food during pregnancy was associated with a protective effect on the appearance of islet autoimmunity in DAISY offspring [111]; however, these findings were not confirmed in a more recent Finnish study (DIPP) [112]. A study in Norway [109] found that cod liver oil taken during pregnancy was associated with reduced risk of T1DM, suggesting that vitamin D and/or the omega-3 fatty acids in the cod liver oil have a protective effect. In the prospective DAISY study, dietary intake of omega-3 fatty acids and the omega-3 fatty acid content of erythrocyte membranes were associated with reduced risk of islet autoimmunity [113] (Figure 2.4). However, neither intake nor membrane levels of omega-3 or omega-6 fatty acids were associated with risk of developing T1DM in those children with islet autoimmunity [114]. A Nutritional Intervention to Prevent T1D (NIP) is currently underway to examine whether nutritional supplements with docosahexaenoic acid (DHA), given during the last trimester of pregnancy and the first few years of life, can prevent development of islet autoantibodies [115].

Gene–environment interaction in clinical type 1 diabetes
T1DM is likely caused by an interactive effect of genetic and environmental factors within a limited age-window. While both the susceptibility genes and the candidate environmental exposures appear to be quite common, the likelihood of these causal components meeting within the susceptibility age-window is low. To investigate the environmental causes of T1DM, the study subjects have to be screened for known susceptibility gene markers so that gene–environment interactions can be accounted for.

Interaction between HLA Class II alleles and viral infection: Susceptibility to diabetogenic enteroviruses in humans appears to be genetically restricted by HLA-DR and DQ alleles [116]. However, the allelic specificity is controversial and may depend on the viral type and epidemicity. In general, the HLA-DR3 allele, present in most patients with T1DM, is associated with viral persistence.

Interaction between HLA Class II alleles and infant diet: Few studies to date have examined a possibility of an interaction between the HLA genes and dietary exposures [117]. The epidemiologic data are limited, but suggest that an early exposure to cow’s milk in relatives with HLA-DR3/4,DQB1*0302, DR3/3 or DRx/4,DQB1*0302 is not associated with development of β-cell autoantibodies [101].
Established type 1 diabetes

Clinical onset
In industrialized countries, 20–40% of T1DM patients younger than 20 years present with diabetic ketoacidosis [118]. After adjusting for age, gender, ethnicity, diabetes type, and family history of diabetes, diabetic ketoacidosis at diagnosis was associated with lower family income, less desirable health insurance coverage, and lower parental education [3]. Younger children present with more severe symptoms at diagnosis, because children younger than 7 years old have lost on average 80% of the islets, compared to 60% in those 7–14 years old and 40% in those older than 14 years [119]. Case fatality in industrialized countries ranges between 0.4–0.9% [120]. Both diabetic ketoacidosis and onset death are largely preventable, because most of the patients have typical symptoms of polyuria, polydipsia, and weight loss 2–4 weeks prior to diagnosis. The diagnosis is straightforward in almost all cases, based on the symptoms, random blood glucose over 200 mg/dL and/or HbA1c >7%.

Traditionally, nearly all children with newly diagnosed T1DM were hospitalized. More recently, an increasing proportion of new-onset children have been managed on an outpatient basis, especially in urban centers with specialized diabetes education and treatment facilities. Hospitalization at onset does not improve short-term outcomes such as re-admission for diabetic ketoacidosis or severe hypoglycemia [118], if adequate family education and follow-up is available on outpatient basis.

Remission (honeymoon period)
Shortly after clinical onset, most T1DM patients experience a transient fall in insulin requirement due to improved β-cell function. Total and partial remissions have been reported in, respectively, 2–12% and 18–62% of young T1DM patients [118]. Older age and less severe initial presentation of T1DM and low or absent ICA or IA-2 [121] have been consistently associated with deeper and longer remission. Evidence relating GAD autoantibodies [121,122], non-Caucasian origin, HLA-DR3 allele, female gender, and family history of T1DM to a less severe presentation, greater frequency of remission, and slower deterioration of insulin secretion is inconclusive. Most studies agree that preserved β-cell function is associated with better glycemic control (lower HbA1c) and preserved α-cell glucagon response to hypoglycemia.

The natural remission is always temporary, ending with a gradual or abrupt increase in exogenous insulin requirements. Destruction of β cells is complete within 3 years of diagnosis in most young children, especially those with the HLA-DR3/4 genotype. It is much slower and often only partial in older patients [123], 15% of whom have still some β-cell function preserved 10 years after diagnosis.

Acute complications
Acute complications of T1DM (diabetic ketoacidosis, hypoglycemia, and infections) are described in detail in other chapters. The risk of hospital admission for acute complication is 30/100 patient-years in the first year of the disease and 20/100 patient-years in the subsequent 3 years [118]. An estimated 26% of the patients have at least one episode of severe hypoglycemia within the initial 4 years of diagnosis. The incidence of severe hypoglycemic episodes varies between 6 and 20 per 100 person-years, and increases with younger age, longer duration of diabetes, intensity of insulin treatment, lower levels of HbA1c, and in older children with presence of underinsurance and psychiatric disorders [118,124]. The incidence of ketoacidosis is about 8 per 100 person-years and increases with age in girls; the risk of ketoacidosis also increases with higher HbA1c, higher reported insulin dose, and in older children with limited access to care due to underinsurance and presence of psychiatric disorders [124]. Interestingly, most of ketoacidosis and/or hypoglycemic episodes occur among 20% of children who have recurrent events.

Morbidity and mortality
Insulin treatment dramatically prolongs survival but it does not cure diabetes. Excess mortality seems to be lowest in Scandinavia, intermediate in the US, and highest in countries where T1DM is rare, for example, Japan, probably due to a combination of the quality of care and access. On the other hand, 40% of the patients survive over 40 years and a half of these have no major complications. Several studies have shown that survival in T1DM has improved over time [125]. The Pittsburgh Epidemiology of Diabetes Complications (EDC) study cohort has recently shown that the life expectancy for those diagnosed 1965–1980 was 15 years greater than participants diagnosed 1950–1964, a difference that persisted regardless of sex or pubertal status at diagnosis [126]. Both the Finnish Diabetic Nephropathy (FinnDiane) study and the Pittsburgh EDC study report that in the absence of renal disease and microalbuminuria, the long-term mortality risk in T1DM is not increased compared with the general population [127].

References


37 Valdes AM, Erlich HA, Noble JA: Human leucocyte antigen class I and II as predictor of Type 1 Diabetes (T1D) susceptibility and age at T1D onset. *Human Immunology* 2005;66(3):301–313.


40 Baschal EE, Aly TA, Babu SR, et al.: HLA-DPB1*0402 protects against type 1 diabetes autoimmunity in the highest risk DR3-DQB1*0201/DR4-DQB1*0302 DAISY population. *Diabetes* 2007;56(9):2405–2409.


43 Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447(7145): 661–678.


75 The Diabetes Control and Complications Trial Research Group: Effect of intensive therapy on residual beta-cell function in patients


84 Stene LC, Rewers M: Immunology in the clinic review series; focus on type 1 diabetes and viruses: the enteroviruses link to type 1 diabetes: critical review of human studies. *Clinical & Experimental Immunology* 2012;168(1):12–23.


CHAPTER 3

Epidemiology and geography of type 2 diabetes mellitus

Qing Qiao1,2, Desmond E. Williams3, Giuseppina Imperatore3, K.M. Venkat Narayan3 and Jaakko Tuomilehto1,2,4

1University of Helsinki, Helsinki, Finland
2National Public Health Institute, Helsinki, Finland
3Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Atlanta, GA, USA
4Danube University of Krems, Krems, Austria; King Abdulaziz University, Jeddah, Saudi Arabia

Age- and sex-specific prevalence of type 2 diabetes in different ethnic groups

Type 2 diabetes mellitus (T2DM) is now taking its place as one of the main threats to human health in the twenty-first century [1]. In 1921, Dr Elliot Joslin was already concerned that according to his count there had been a doubling of diabetest in three decades [2]. The impact of T2DM is increasingly felt around the world, with its prevalence rising dramatically over recent decades. The World Health Organization (WHO) estimated that there were 150 million people aged 20 years and older living with diabetes in 2000, and by 2025, this will have risen to 300 million (Figure 3.1). There will be a 42% increase, from 51 to 72 million, in developed countries and a 170% increase, from 84 to 228 million, in developing countries [3]. The top 10 countries with the highest estimated number of people with diabetes in 2025 are listed in Table 3.1.

Diagnostic criteria for diabetes

In the past two decades there have been several important developments, which have had significant impact on the definition of diabetes and thereby on the assessment of its magnitude. In 1979, the 2-h 75-g oral glucose tolerance test (OGTT) was proposed as a standard test for diagnosis of diabetes by the National Diabetes Data Group (NDDG) [4], and endorsed by WHO in 1980 [5]. In 1985, WHO made a minor modification to the diagnostic criteria [6]. This has created order out of the confusion in the diagnostic criteria for diabetes. Before that enormous variations existed in diagnostic cutoff values for fasting as well as after glucose loading. The glucose load varied between 50 and 100 g or was body weight related. The differences in glucose assay methods, glucose load, and the time after loading made the comparison between different studies difficult. Near-universal adoption of the WHO criteria has had a significant influence on epidemiologic studies of diabetes.

In 1997, a revision of the diagnostic criteria was approved by the American Diabetes Association (ADA) [7] and adopted by WHO Consultation in 1999 [8]. The major changes were lowering the positive cutoff value of fasting venous plasma glucose from 7.8 mmol L−1 (140 mg dL−1) to 7.0 mmol L−1 (126 mg dL−1). The positive cutoff value for 2-h plasma glucose remained unchanged, that is, 11.1 mmol L−1 (200 mg dL−1) and over. For epidemiologic studies and for routine clinical practice, the ADA did not recommend the primary use of OGTT, but WHO Consultation still retained the OGTT as the standard test procedure. The prevalence data assembled in this chapter are estimated mainly according to the WHO 1999 criteria [8], except for those noted otherwise.

Conversion factors for glucose concentrations

Glucose concentration can be measured using different blood specimens such as venous plasma glucose, capillary whole blood glucose, and venous whole blood glucose. The cutoff points for the classification of stages of glucose abnormalities from different specimens are different. Currently, there are no internationally accepted conversion factors for glucose concentrations in the literature. Recently, conversion factors for changing glucose concentrations between different blood samples were developed on the basis of data from a Finnish study and applied them in the DECODE study (Diabetes Epidemiology: Collaborative Analysis of Diagnostic Criteria in Europe) [9–11]. The equations were derived based on 294 matched samples of whole blood (capillary and serum) glucose and plasma glucose concentrations drawn from a standard 75-g OGTT in 74 individuals at 0, 30, 60, and 120 min at the
Chapter 3

Figure 3.1 Number of people with diabetes mellitus, 1995–2025. Source: Adapted from King et al. [3].

Table 3.1 Top 10 countries for estimated number of adults with diabetes, 1995 and 2025

<table>
<thead>
<tr>
<th>Rank</th>
<th>Country</th>
<th>1995 (millions)</th>
<th>Country</th>
<th>2025 (millions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>India</td>
<td>19.4</td>
<td>India</td>
<td>57.2</td>
</tr>
<tr>
<td>2</td>
<td>China</td>
<td>16.0</td>
<td>China</td>
<td>37.6</td>
</tr>
<tr>
<td>3</td>
<td>United States</td>
<td>13.9</td>
<td>U.S.</td>
<td>21.9</td>
</tr>
<tr>
<td>4</td>
<td>Russian Federation</td>
<td>8.9</td>
<td>Pakistan</td>
<td>14.5</td>
</tr>
<tr>
<td>5</td>
<td>Japan</td>
<td>6.3</td>
<td>Indonesia</td>
<td>12.4</td>
</tr>
<tr>
<td>6</td>
<td>Brazil</td>
<td>4.9</td>
<td>Russian Federation</td>
<td>12.2</td>
</tr>
<tr>
<td>7</td>
<td>Indonesia</td>
<td>4.5</td>
<td>Mexico</td>
<td>11.7</td>
</tr>
<tr>
<td>8</td>
<td>Pakistan</td>
<td>4.3</td>
<td>Brazil</td>
<td>11.6</td>
</tr>
<tr>
<td>9</td>
<td>Mexico</td>
<td>3.8</td>
<td>Egypt</td>
<td>8.8</td>
</tr>
<tr>
<td>10</td>
<td>Ukraine</td>
<td>3.6</td>
<td>Japan</td>
<td>8.5</td>
</tr>
<tr>
<td></td>
<td>All other countries</td>
<td>49.7</td>
<td>Japan</td>
<td>103.6</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>135.3</td>
<td>300.0</td>
<td></td>
</tr>
</tbody>
</table>

Source: Adapted from King et al. [3].

Diabetes and Genetic Epidemiology Unit, National Public Health Institute in Finland. The relationships between glucose concentrations as measured by the different methods used were estimated. The formulas derived are as follows:

Venous plasma glucose (mmol L\(^{-1}\))

\[
= 0.558 + 1.119 \times \text{whole blood glucose (mmol L}^{-1}\)\]

Venous plasma glucose (mmol L\(^{-1}\))

\[
= 0.102 + 1.066 \times \text{capillary blood glucose (mmol L}^{-1}\)\]

Venous plasma glucose (mmol L\(^{-1}\))

\[
= -0.137 + 1.047 \times \text{serum glucose (mmol L}^{-1}\)\]

**Age- and sex-specific plasma glucose concentration**

The age- and sex-specific mean fasting and 2-h plasma glucose after 75-g glucose load were estimated in general Caucasian populations in Europe, and in Chinese, Japanese, and Indians in Asia, who did not have prior history of diabetes. The participants are included in the DECODE and the DECODA (Diabetes Epidemiology: Collaborative Analysis of Diagnostic Criteria in Asia) studies, the two largest epidemiologic studies for diabetes in Europe and Asia, with a total of 15,660 subjects from Europe [10] and 19,845 subjects from Asia [12]. The data presented here are based on the pooled data from 13 DECODE and the 11 DECODA participating cohorts. A standard OGTT was carried out in all populations, and subjects with prior history of diabetes were not included in the analysis. The plasma glucose concentrations rose with age and reached a peak at 60–69 years of age and then started to decline in Indians but continued to increase after 70 years of age in Europeans (Figure 3.2). In each age group, the mean 2-h plasma glucose was significantly higher for Indians than for Chinese and Japanese, and the same was also true for fasting plasma glucose in most of the age groups (Figure 3.2). The mean fasting and 2-h glucose concentrations did not differ between Chinese and Japanese except at 40–49 years of age where the glucose values were higher in the Japanese. The mean glucose levels were lower in Europeans than in Asians younger than 70 years, whereas they were higher in Europeans than in Asians 70 years or older. The mean glucose levels were similar in Asian men and women. In Europe, the mean fasting glucose concentration was higher in men than in women at 30–69 years of age but after 70 years of age, it was higher in women. The 2-h glucose was higher in women than in men throughout the age range. Two-hour glucose increased more with age than did fasting glucose.

**Age- and sex-specific prevalence of diabetes**

Europe

The prevalence of diabetes has been estimated by applying the WHO 1999 criteria [8] for 13 European and 11 Asian cohorts participating in the DECODE and the DECODA studies. In Europe, the age-specific prevalence of diabetes rose with age up to 70s and 80s in both men and women [10] (Figure 3.3). In most of the studies, the prevalence was less than 10% in subjects younger than 60 years and between 10 and 20% at 60–79 years of age. They were higher in Malta than in other populations. The prevalence of isolated postload hyperglycemia (2-h glucose ≥11.1 mmol L\(^{-1}\) and fasting glucose <7.0 mmol L\(^{-1}\)) increased more with age than did isolated fasting hyperglycemia (fasting glucose ≥7.0 mmol L\(^{-1}\) and 2-h glucose <11.1 mmol L\(^{-1}\)), particularly in women.

A recent German study using WHO 1999 criteria [8] showed that the prevalence of diabetes in Germany in 2000 was 16.7% in men and 8.6% in women at 55–59 years of age and 23.1% in men and 17.0% in women at 70–74 years of age [13]; that is, rates are within the variation reported for the DECODE study: In the recent Turkish Diabetes Epidemiology Study (TURDEP-II) undertaken in 2010, the prevalence of diabetes was 16.5% and increased with age, reaching a peak of 37.7% in urban men at 70–74 years and 43.6% in urban women.
Figure 3.2 Mean fasting (the two lower lines) and 2-h (the two upper lines) plasma glucose concentrations and their 95% confidence intervals (vertical bars) for the DECODE and the DECODA study populations. * Indians vs. Chinese, + Indians vs. Japanese, § Japanese vs. Chinese, all for $p<0.05$ for both genders combined. Source: Adapted from the DECODE [10] and the DECODA [12] Study Groups.

at 75–79 years of age [14]. These rates are higher than those reported by an earlier Turkish study in 1997–1998 [15].

Compared with most of the other racial and ethnic groups worldwide, where age- and sex-specific prevalence of diabetes has been reported, Europeans have a moderate to low prevalence of diabetes [10].

United States

In the United States, the prevalence of diabetes varies considerably among different ethnic groups. The prevalence was 1.9 times greater in Latino Americans and 1.6 times in African Americans than in Whites of the same age in the Third National Health and Nutrition Examination Survey (NHANES III) [16]. In NHANES III, where a single fasting plasma glucose $\geq 7.0 \text{mmol L}^{-1}$ was applied for the diagnosis of diabetes, the prevalence of diabetes in US Whites in 1988–1994 was 5.9% in men and 4.8% in women at 40–49 years of age and reached a peak of 19.2% in men and 16.6% in women at 75 years or older [16]. The prevalence of undiagnosed diabetes according to the same fasting glucose criteria at a comparable age range of 40–59 years was higher in US Whites than in most of the female and in half of the male European populations participating in the DECODE study [10]. The rates were higher in Mexican Americans than in US Whites, and were higher than seen in any of the populations included in the DECODE study.

The Pima Indians have the highest prevalence of diabetes in the world, being 50% at 30–64 years of age, estimated using 2-h postload glucose test alone [17]. The prevalence of diabetes in Native Hawaiians (Polynesian population of Hawaii) in two rural communities was estimated using WHO 1985 criteria [18]. At 30–39 years of age, the prevalence was 6.5% in women and 10.7% in men, reaching a peak of 34.6% in women and 40.0% in men at 70 years or older. Other Native American tribes also have a higher prevalence of diabetes than Caucasoids do.

Central and South America

The age-standardized prevalence of diabetes using 2-h glucose criteria alone was investigated in a Brazilian population in Sao Paulo in 1987 and in a Colombian population in Bogota in 1988–1989 [17]. The prevalences in both populations were similar, around 7% for men and 9% for women. The age-specific prevalence of diabetes in a Mexican population in Mexico City
was 4.2% in men and 3.2% in women at 35–39 years of age and reached 23.1% in men and 41.7% in women at 60–64 years of age [17].

**Australia**
The Australia Diabetes, Obesity and Lifestyle Study (AusDiab) took place in 1999–2000 applying WHO 1999 criteria mainly in Caucasoids. The prevalence of diabetes increased with age, from 2.7% in men and 2.2% in women at 35–44 years of age to 23.5% and 22.7% at 75 years or older [19]. These rates were higher than those in most of the DECODE populations [10].

**Asia and Pacific Islands**

**Asia** The prevalence of diabetes varies markedly among Asian populations. In those participating in the DECODA study, it rises with age up to 70s and 80s in Chinese and Japanese, and in Indian men and women with age up to 60s and then declines [12] (Figure 3.4a,b). The age- and sex-specific prevalence and the peak prevalence of diabetes were higher in cohorts from India and Singapore than in most of the Chinese and Japanese cohorts. In Chinese and Japanese, the prevalence was less than 10% at 30–49 years of age; the peak prevalence was less than 20% in most of the cohorts and none exceeded 30%. In contrast, in India and Singapore the prevalence was over 10% among people aged 40–49 years, and over 30% among those aged 50–69 years for most of the cohorts. The urban Chinese and Japanese had significantly higher prevalences of diabetes than their rural counterparts at 40–69 years of age in men and at 50–79 years of age in women (Figure 3.4). The prevalence of diabetes in Korea was within the range observed in China and Japan [20].

Type 2 diabetes was found at a relatively younger age in Pakistan and the prevalence reached the peak in the age group 55–64 years [21–23]; the prevalence pattern was similar to that in India. In a rural community the prevalence was over 13% at 35–44 years of age, with the highest prevalence of 30% in men at 65–74 years of age [22], indicating diabetes has already become a major health threat in Pakistan as in India.

In the late 1990s, King et al. carried out a series of surveys on diabetes in Uzbekistan and Mongolia using WHO 1985 criteria [6]. In Uzbekistan, the prevalence of T2DM was relatively low in people younger than 45 years, around 1%, and 10–15% at 65 years or older [24]. It was relatively rare in people younger than 45 years, around 1%, and 10–15% at 65 years or older. The prevalence in Mongolia [25] was relatively low with a peak of less than 5% at the age of 65 years or older, comparable to the prevalence reported from China in 1994 [26]. Taking into account the high positive cutoff value of 7.8 mmol L$^{-1}$ for fasting glucose for diabetes in the WHO 1985 criteria, the prevalence reported in these studies would be somewhat higher if the current cutoff value of 7.0 mmol L$^{-1}$ were used. Because different diagnostic criteria have been applied, the results from some of these studies cannot be compared directly with those from the DECODA study [12].

Compared with the European populations included in the DECODE study [10], the age-specific prevalence of diabetes in urban Chinese and Japanese was slightly higher than that in Europeans at 30–69 years of age, but was lower than that in Indians. In the elderly population, however, the peak prevalence was higher in a few European populations than in Indians, such as in Maltese, in Finnish women in Oulu, and in women living in the Canary Islands, Spain. The age at which the peak prevalence
Figure 3.4 Age- and sex-specific prevalence of diabetes in Asia men (a) and women (b) in the DECODA study. DMF, diabetes determined by FPG $\geq$ 7.0 mmol L$^{-1}$ and 2hPG $<$ 11.1 mmol L$^{-1}$; DMP, diabetes determined by 2hPG $\geq$ 11.1 mmol L$^{-1}$ and FPG $<$ 7.0 mmol L$^{-1}$; DMF&DMP, diabetes determined by FPG $\geq$ 7.0 mmol L$^{-1}$ and 2hPG $\geq$ 11.1 mmol L$^{-1}$.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, for the difference between urban and rural Chinese and Japanese; + $p < 0.05$, +++ $p < 0.001$, for the difference between Chinese and Japanese combined and Indians. Source: Adapted from the DECODA Study Group [12].
of diabetes was reached was similar for Europeans, Chinese, and Japanese (over 70 years), while Indians had their highest prevalence at the age up to 60s and then started to decline. These differences in prevalence in the elderly (women, in particular) are probably due to selective mortality associated with diabetes. The recent survey in China showed a marked increase in the prevalence of T2DM; it was 9.7%, representing an estimated 92.4 million adults in China with diabetes in 2007 [27].

In all Asian populations included in the DECODA study, the prevalence of isolated fasting hyperglycemia (fasting plasma glucose ≥7.0 mmol L⁻¹ and 2-h plasma glucose <11.1 mmol L⁻¹) did not increase with age (Figure 3.4). The prevalence of isolated 2-h hyperglycemia (2-h plasma glucose ≥11.1 mmol L⁻¹ and fasting plasma glucose <7.0 mmol L⁻¹) tended to increase with age in Chinese and Japanese but not in Asian Indians [12].

Pacific Islands There are remarkable differences in the prevalence of diabetes among the Melanesia, Micronesian, and Polynesian populations of the Pacific Islands. According to the 2-h postload glucose criteria alone [6], an age-standardized diabetes prevalence of more than 40% was revealed in the Micronesia population of Nauru in the 1980s [28,29]. The prevalence of diabetes in the Melanesia population of Papua New Guinea had been reported close to 0% in highland populations [30], whereas in the urbanized Koki people the age-standardized rate exceeds 40%, approaching that of Nauru [29], exhibiting an extreme urban–rural gradient. Intermediate rates are seen in other Pacific Island populations. In the Polynesia population of the Western Samoa, the crude prevalence rates were 3.4 and 8.7% in rural and urban populations, respectively. By 1991, these rates had risen to 6.5 and 9.0% in two rural communities and to 16% in the urban settings of Apia [31]. A recent study in the Polynesia population of Tonga in 1998–2000 using 2-h OGTT showed that the peak prevalence of diabetes reached 20% in men and 40% in women aged 60 years or older [32].

Middle East The prevalences of diabetes in Arabian countries, calculated according to the WHO 1985 criteria [6], have been reported to be high [33–35]. It is relatively low before the age of 30 years and starts to increase during the 40s, with the highest in the oldest age group (Figure 3.5). In a rural Palestinian village, in 1996 the prevalence of diabetes was less than 4.0% in people younger than 40 years but increased to 11.0% at 40–49 years of age, with a peak of 21.7% in men and 31.6% in women at 60–65 years of age [35]. The prevalence in Palestine was similar to that in rural Saudi Arabia, but both were lower than that in urban Saudi Arabia [33]. The prevalence of T2DM in 1995–1996 in Kuwait was recalculated using ADA 1997 criteria [7]. It was lower than 3% at age 20–29 years, around 9% at age 30–39 years and higher than 15% at age 40–49 [36]. Diabetes is prevalent in all Arab countries despite the differences in economic status among these countries, indicating that genetic susceptibility and cultural factors may play an important role in the development of the disease.

The age-specific rate ranged from 8% at 40–44 years of age to 25% at 60–64 years of age in Israeli Jews [17].

Africa In subjects aged 30–64 years, the age-standardized prevalence of diabetes using 2-h glucose alone has been reported to be higher in Hindu and Muslim Indians living in Mauritius [37] and Tanzania [38], around 10% in Tanzania and 13–18% in Mauritius. The age-standardized rate was very low in Bantu in Tanzania, less than 1% in women and 0.9–3.3% in men [39]. The prevalence was 8% in Tunisia [17]. It is interesting to note that the age-standardized prevalence was much higher in

![Figure 3.5 Prevalence of diabetes in men and women in three Arabian countries. (For a color version of this figure, please see the color plate section.)](image-url)
Chinese living in Mauritius than that in Da Qing in China in the mid 1980s [17,37], indicating the importance of impact of the environmental factors.

In Native People: Mapuche and Aymara in Chile and in Siberia in Russian Federation
The native people who still practice their traditional lifestyle and undertake considerable physical activity have extremely low prevalences of diabetes despite their high prevalence of obesity [40,41]. Among Aymara the prevalence of diabetes in 1997 was almost undetectable despite the fact that 13% of the men and 24% of the women had a body mass index (BMI) higher than 30 kg m$^{-2}$ [40]. Similar findings have been reported among native Mapuche [41,42]. The indigenous groups in northern Siberia also showed a very low prevalence, being less than 1% in 1994 [43]. This suggests that a healthy lifestyle with much physical activity provides protection from the development of diabetes. A recent report from Mapuche [41] showed that the prevalence in 1998 was higher than that reported 15 years ago, as was the prevalence of obesity [42], suggesting a possible impact of lifestyle changes on the trends in prevalence of diabetes.

Previously undiagnosed diabetes
The proportion with previously undiagnosed diabetes varies with age. It seems to be highest at 30–39 years of age (70–80%), and lowest in the elderly (around 40%) [10,12]. The only exception was seen in European women where the proportion of undiagnosed diabetic cases was around 40–45% in all age groups. The proportion of undiagnosed diabetes was higher in European men than in European women. In Asians, it was slightly higher in women than in men in the youngest age groups.

Prevalence of impaired glucose tolerance and impaired fasting glycaemia in different ethnic groups
A category of nondiabetic fasting hyperglycaemia was defined only recently, by the ADA in 1997 [7] and adopted also by WHO in 1999 [8], and named impaired fasting glycaemia (IFG). It was introduced by consensus to define impaired glucose homeostasis intermediate between diabetes and normal glucose homeostasis and to be analogous to impaired glucose tolerance (IGT), but without any epidemiologic evidence of possible risks associated with it. Since the introduction of the category of IFG, prospective studies have examined the relationship between IFG and future morbidity and mortality with a comparison to IGT, and shown that the risk of cardiovascular disease (CVD) morbidity and mortality is higher for IGT than for IFG [44–47]. Thus far the data are scarce on the risk of progression to diabetes in subjects with IFG as compared with those with normal fasting glucose or those with IGT. A few studies, which have examined the issue, agree that the risk of developing diabetes is high in subjects with either IFG or IGT and highest in those with both IFG and IGT, as compared with subjects with normal fasting and normal 2-h glucose [48–52]. At present neither IFG nor IGT is considered a clinical entity, but as a risk category for the future development of diabetes [53]. Each represents a metabolic state intermediate between normal glucose homeostasis and diabetic hyperglycaemia, and they were combined and defined formally as impaired glucose regulation (IGR) by WHO Consultation in 1999 [8]. All studies agree that only IGT but not IFG is a risk factor for CVD.

The prevalences of IGT and IFG in Europe and Asia were reported recently by the DECODE and the DECODA Study Groups [8,12].

Europe
The prevalence of IGR rose with age in each study [10] (Figure 3.6). In most of the study populations, the prevalence of IGR was less than 15% at 30–59 years of age and between 15 and 30% after 60 years of age. The prevalence of IGT increased linearly with age, but the prevalence of IFG did not (Figure 3.6). The increase in the prevalence of undiagnosed diabetes and IGR in the elderly population resulted mainly from the proportionately larger increase in postload hyperglycaemia than in fasting hyperglycaemia.

Asia
The prevalence of IGR rose with age up to the 70s and 80s in most of the study cohorts [12] (Figure 3.7a,b). The increase was graded with aging in Chinese, Japanese, and Singaporean populations, as observed in Europeans, but not in Indians where the prevalence of IGR started to increase by the age of 30–39 years and did not change much with increasing age. The peak prevalences of IGR were not different among different populations, but the age-specific prevalence of IGR was higher in Indians than in Chinese and Japanese at 30–49 years of age for both men and women. In the urban populations the prevalence was higher than in the rural populations aged 40–69 years in men and 50–59 years in women in the Chinese and Japanese populations (Figure 3.7). The difference in the prevalence pattern in different ethnic groups may not be completely explained by living environments and geographic locations, suggesting that genetic differences also play a role.

IGT was more prevalent than IFG in almost all age groups in Asian subjects (Figure 3.7). The prevalence of IGT increases with age whereas IFG does not. This pattern is consistent with that among the European populations [10]. The concordance for IFG and IGT was very poor in all populations, particularly in Asians [10,12,54–56]. The finding that postload hyperglycaemia was more prevalent in the elderly in Europe and Asia is consistent with the report from NHANES III [57]. Thus, the prevalence of undiagnosed diabetes and IGR would be underestimated to a large extent, especially in female and elderly populations, if only fasting glucose determination were used. The primary purpose of population-based testing for blood glucose is to detect previously undiagnosed diabetes and IGR in order to apply early intervention to reduce the serious
diabetic complications and to prevent progression from IGT to diabetes as demonstrated by the recent diabetes prevention trials [58–62].

**Sex differences in prevalence of diabetes, IGT, and IFG**

The ratio of prevalences of glucose abnormality between men and women has been estimated in many studies, but so far there has been no consistent trend [16,17]. In the DECODE study, we found there is a clear pattern in the prevalence of postload hyperglycemia and the prevalence of fasting hyperglycemia by sex [10]. Undiagnosed diabetes and IFG defined by isolated fasting hyperglycemia was more common in men than in women at 30–69 years of age, whereas the prevalence of isolated postload hyperglycemia was higher in women than in men and was particularly high in the elderly population [10]. In the DECODA study, sex difference was not as clear as in Europe. The prevalence of IFG also seems higher in Chinese and Japanese men than in women, whereas it was higher in Indian women than in men. IGT was more prevalent in Chinese and Japanese women than in men, but such a difference was not observed in Indians [12]. Sex differences in the prevalence of diabetes and IGR depend on how the prevalence was estimated, by fasting or by postload hyperglycemia, on the age distributions, and on the ethnic groups. Asian Indians, who have a very high risk of diabetes, show abnormalities in fasting glucose values at an earlier age than other populations.

**The ratio of IGT to diabetes**

The ratio of IGT to diabetes has been reported to decrease as prevalence of diabetes rises [17] and may have some predictive value in determining the stage of a glucose intolerance epidemic within a population [63]. When the ratio is high but the prevalence of diabetes is low, the early stage of a diabetes epidemic may be occurring [17]. The age- and sex-specific ratios of IGR to diabetes according to the newly revised diagnostic criteria for diabetes [8] are shown in Figure 3.8a for Asian and Figure 3.8b for European populations [10,12]. The ratio of IGR to diabetes declined when the prevalence of diabetes increased in both Asian and European populations.

**Secular trends in prevalence of type 2 diabetes**

Accumulating evidence shows that the prevalence of diabetes is increasing with time over recent decades. This upward trend has been seen primarily in developing countries [3,64] (Figure 3.1). A series of studies in the southern Indian city of Chennai showed a steady increase in the prevalence of diabetes in the
Figure 3.7  Age- and sex-specific prevalence of impaired glucose regulation in the DECODA study populations in men (a) and women (b). Isolated-IFG, FPG 6.1–6.9 mmol L\(^{-1}\) and 2hPG <7.8 mmol L\(^{-1}\); Isolated-IGT, 2hPG 7.8–11.0 mmol L\(^{-1}\) and FPG <6.1 mmol L\(^{-1}\); IFG&IGT, FPG 6.1–6.9 mmol L\(^{-1}\) and 2hPG 7.8–11.0 mmol L\(^{-1}\). *p <0.05, **p <0.01, ***p <0.001, for the difference between urban and rural Chinese and Japanese; + p <0.05, +++ p <0.001, for the difference between Chinese and Japanese combined and Asian Indian populations combined. Source: Adapted from the DECODA Study Group [12].
Indian population. The age-standardized prevalence increased from 8.2% in 1988–1989 [65] to 11.6% in 1994–1995 [66] and reached 13.5% in 2000 [67], a 65% increase within the two decades. During the last two decades a considerable amount of information has been obtained from China. Studies conducted in China between 1980 and 1990 consistently show low diabetes prevalence rates of approximately 1.5% or less, even in urban populations such as in Shanghai in 1980 [68–71]. The prevalence of diabetes in Shanghai in 1980 was close to 1%. In rural Guangdong province it was 0.33% [69]. Studies undertaken in the late 1990s, however, indicate sharply rising prevalence rates in China [72–74] and the rates estimated at the beginning of the current century show that diabetes in an urban Chinese population in mainland China [75] is already as prevalent as in Hong Kong and Taiwan in the mid 1990s (Figure 3.9) [76–78]. In 2007, the prevalence of T2DM in China was almost 10% indicating a three-fold increase in three decades [27]. In Turkey, the prevalence of T2DM doubled during a 12-year interval from 1998 to 2010 [24]. The prevalence of both diabetes and its microvascular complications in a Pacific Island population (20 years or older) of Western Samoa was examined in 1978 and 1991 [31]. In 1978, the crude prevalence rates were 3.4 and

Figure 3.8 Age- and sex-specific scatter plot of ratios of IGR to diabetes against the prevalence of diabetes (diagnosed and undiagnosed) separately for 11 Asian (a) and 13 European (b) studies. Source: Adapted from the DECODE [10] and the DECODA [12] Study Groups.
Figure 3.9 Age-standardized prevalence of type 2 diabetes mellitus in a series of nationwide surveys in China and a recent study in Qingdao city (personal communication with Weiguo Gao).

8.7% in rural and urban populations, respectively. By 1991, these rates had risen to 6.5 and 9.0% in two rural communities and to 16.0% in the urban setting of Apia.

Recent studies indicate that diabetes prevalence continues to increase even in developed countries. According to the data from the US NHANES II and NHANES III surveys, using ADA criteria, the prevalence of T2DM in the US adult population aged 40–74 years of age increased from 8.9% in the period 1976–1980 to 12.3% by 1988–1994. A similar increase was found when WHO criteria were applied (11.4 and 14.3%) [16].

In Australia, the total prevalence of diabetes had increased from 3.4 to 7.2% from 1981 to 1999–2000 and the difference persisted after adjustment for BMI [19] (Figure 3.10). In an adult Norwegian population [79], the crude prevalence of known diabetes in men increased from 2.6% in 1984–1986 to 3.3% in 1995–1997, an increase of 24%, but the increase was not found in women. Over the same time period, an increase of 86% in the prevalence of obesity defined by BMI ≥30 kg m\(^{-2}\) was observed in men, which was much higher than the increase of 38% in women. In a Danish study of a 60-year-old cohort over a 22-year period an increase of 58% in men and 21% in women in the prevalence of diabetes was observed, which was fully explained by a concurrent increase in BMI [80]. Rising trends in the prevalence of diabetes and obesity have also been reported in other European countries [81,82]. In addition to the increase in obesity, reduced physical activity resulting from changes in work-related activity and sedentary lifestyle has contributed to the rising trend in T2DM.

Prevalence, which reflects the accumulation of the patients at any given time, can be influenced by many factors such as an increase in the number of new cases and a reduction in the mortality attributed to the disease. There is evidence that mortality in diabetes has declined in men in the United States [83]. Thus, a rise in prevalence could be a result of an improved survival of diabetic subjects. However, studies also show an increasing trend in diabetes incidence due to the increase in obesity and decrease in exercise. Knowler et al. [84] compared the incidence rates over two 10-year time periods, 1965–1975 and 1975–1985, in Arizona Pima Indians, and found that over the 10-year period the incidence rates increased by 50% in most age and sex groups. The San Antonio Heart Study revealed an increasing secular trend in the 7- to 8-year incidence of T2DM occurring from 1987 to 1996 in Mexican American and non-Hispanic Whites [85]. Therefore, both increased incidence and decreased mortality among diabetic subjects have contributed to the increased trend in the prevalence of diabetes.

Type 2 diabetes in children

Prevalence and incidence
Type 2 diabetes was historically a rare occurrence in children but recent studies have reported marked increases in the prevalence of T2DM in children. Type 2 diabetes was first reported in a population-based study in 1979 of American Indians children in Arizona [86]. This American Indian community has one of the highest rates of T2DM in adults and obesity in both adults and children [87]. After 30 years of follow-up in this population, the youngest age of onset of T2DM was 4 years and the prevalence of T2DM in 15–19-year-old children increased from 2.4 to 3.8% in boys and from 2.7 to 5.3% in girls [88]. Data from the Indian Health Service (IHS) in the United States confirmed an increase in the prevalence of diabetes in American Indian populations in the United States, with a 68% increase in the prevalence of diagnosed diabetes in American Indian and Alaska Natives adolescents between the ages of 15 and 19 years between 1990 and 1998. Although these IHS estimates were for all diabetes and not only T2DM, T1DM is very rare in some of these American Indian populations [89,90].

The increase in T2DM is not limited to the American Indians. Very few population-based studies have been conducted in other racial and ethnic groups but results from diabetes
### Table 3.2 Selected studies of type 2 diabetes in children and adolescents

<table>
<thead>
<tr>
<th>Study and reference*</th>
<th>Years</th>
<th>Race/ethnicity</th>
<th>Age (years)</th>
<th>Sample size</th>
<th>No. of cases</th>
<th>Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Population-based studies</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Mexico, USA [136]</td>
<td>1991–1992</td>
<td>Navajo Indians</td>
<td>12–19</td>
<td>142</td>
<td>2</td>
<td>14.1 [0–33.5]**</td>
</tr>
<tr>
<td>NHANES III, USA [138]</td>
<td>1988–1994</td>
<td>Whites, and African and Mexican Americans, all</td>
<td>12–19</td>
<td>2867</td>
<td>13*</td>
<td>4.1 [0–8.6]**</td>
</tr>
<tr>
<td><strong>Clinic-based studies</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indian Health Services, USA [89]</td>
<td>1996</td>
<td>American Indians, all Americans</td>
<td>0–14</td>
<td>402,580</td>
<td>518*</td>
<td>1.3*</td>
</tr>
<tr>
<td>United Kingdom [100]</td>
<td>1993–2000</td>
<td>British</td>
<td>0–18</td>
<td>261,811</td>
<td>21</td>
<td>0.038</td>
</tr>
<tr>
<td>Clinic-based study§ Libya [96]</td>
<td>1981–1990</td>
<td>Libyan</td>
<td>0–14</td>
<td>11,846</td>
<td>12</td>
<td>0.5 [0.26–0.87]</td>
</tr>
<tr>
<td><strong>Cincinnati, OH, USA [92]</strong></td>
<td>1994</td>
<td>Whites and African Americans</td>
<td>0–19</td>
<td>U</td>
<td>U</td>
<td>3.5</td>
</tr>
<tr>
<td><strong>Case series¶</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>San Diego, CA, USA [105]</td>
<td>1993–1994</td>
<td>Whites, Hispanics, and African and Asian Americans</td>
<td>0–16</td>
<td>160*</td>
<td>13*</td>
<td>8*</td>
</tr>
<tr>
<td>San Antonio, TX, USA [106]</td>
<td>1990–1997</td>
<td>Whites, Hispanics</td>
<td>U</td>
<td>560*</td>
<td>101*</td>
<td>18*</td>
</tr>
<tr>
<td>Ventura, CA, USA [103]</td>
<td>1990–1994</td>
<td>Hispanics</td>
<td>0–17</td>
<td>31*</td>
<td>14*</td>
<td>45*</td>
</tr>
<tr>
<td>Thailand [97]</td>
<td>1996–1999</td>
<td>Thai</td>
<td>0–14</td>
<td>39</td>
<td>7</td>
<td>17.9</td>
</tr>
</tbody>
</table>

Note: Numbers in italics are estimates. NHANES III = National Health and Nutrition Examination Survey III; U = unknown data.

* Table updated from Fagot-Campagna et al. [112].

† Estimates are given as prevalence per 1000 and 95% CI.

‡ Estimates include cases of type 1 diabetes.

§ Estimates are given as incidence per 100 000/year.

¶ Estimates are given as percentage of type 2 diabetes among new cases of diabetes.

†† In case series, the sample size refers to the total number of cases of diabetes (types 1 and 2).

 registries, case reports, and cross-sectional studies from Canada [91], Cincinnati [92], Japan [93–95], Libya [96], Thailand [97], United Kingdom [98–100], India [101], Taiwan [102] and numerous case reports from the United States [103–107] have all indicated a significant increase in the prevalence of T2DM in children, although all these studies did not use a standard method to differentiate between type 1 and type 2 diabetes (see Table 3.2). The incidence of T2DM in African American adolescents in Cincinnati increased 10-fold from 0.7 to 7.2 per 100,000 over a 12-year period [92]. The proportion of T2DM in children diagnosed with diabetes has increased from 2–4% in 1992 to 8–45% in the last decade [108]. The proportion of type 1 to type 2 diabetic children is highly variable according to their different age and racial and ethnic groups (see Table 3.2).
Characteristics and diagnosis
The majority of children with T2DM are non-Caucasian and in one review about 94% of cases were from minority groups [109]. The cause of this increased risk in non-Caucasian groups is unclear but will involve genetic predisposition and cultural and environmental risk factors. African American [110,111] and Hispanic [112] children, for example, have been shown to have higher insulin resistance than do non-Hispanic White children.

The majority of T2DM cases occur in over-weight and obese children and they may have clinical signs of insulin resistance, such as acanthosis nigricans. The prevalence of T2DM in these children is higher in girls who may be two to six times more likely to have T2DM. The mean age of onset is around puberty in most populations.

Cases may present with classical signs and symptoms of diabetes but the disease may have an insidious onset and may only be detected by opportunistic screening. In a series from a referral center in Cincinnati, 32% of children with T2DM were discovered by opportunistic screening [92]. Therefore, underreporting of cases of T2DM in children may be as common as in adults.

The differentiation of T2DM from T1DM may be difficult in some cases (see Table 3.3). Currently, the mainstay of differentiating diabetes in children is inadequate and includes the use of clinical characteristics such as obesity, severity of onset, use of insulin, age of onset, diabetic ketoacidosis, and family history of diabetes. With the rising prevalence of obesity in children, more T1DM cases are presenting with obesity [113]. In addition, some cases of T2DM may present with diabetic ketoacidosis [114,115]. Diabetes-related autoantibodies that include the glutamic acid decarboxylase (GAD), tyrosine phosphatase-like molecule (IA2), islet cell antibody (ICA), and C-peptide (which is co-secreted with insulin) concentrations are currently being investigated as a means of improving the differentiation of diabetes type in children. The use of diabetes autoantibodies is limited by problems with assay methodology and the transient nature of some of these autoantibodies in cases of T1DM [116,117]. The use of C-peptide to differentiate the type of diabetes in children is made especially difficult by varying levels of residual β-cell function in type 1 diabetic cases. Current efforts are underway in a CDC (Centers for Disease Control) and NIDDK (National Institute of Diabetes and Digestive and Kidney Diseases) sponsored study, the SEARCH for Diabetes in Youth Study, to develop cutoff points and algorithms that include diabetes autoantibodies and C-peptide concentration which may be used to diagnose and differentiate diabetes in children.

Risk factors
Very few studies have examined the risk factors for T2DM in children. Some of the established risk factors for adult T2DM, discussed in other sections of this chapter, may play a role in the development of T2DM in children. Overweight and obesity are some of the strongest risk factors for insulin resistance and T2DM in adults. The rising prevalence of overweight in children may thus play a role in the risk of T2DM in children. In the United States, the prevalence of overweight in children has increased sharply over the last three decades to 15% in a 1999–2000 survey [118]. Over 20% of Hispanic and African American children are overweight compared to 12% in non-Hispanic White children [119]. This difference in prevalence of obesity may in part explain the higher prevalence of T2DM in ethnic minority children such as American Indians, African Americans, and Hispanics in the United States [88,92,103] and Asian Indians in the United Kingdom [100].

Most cases of T2DM occur around puberty. The reason for this increased risk around puberty is unclear but may be related to the physiologic 30% increase in insulin resistance observed in children as they go through Tanner stages II to IV [120–124]. This physiologic increase in insulin resistance is thought to be related to increased growth hormone production during puberty [121,122] but not related to the increase of sex hormones [121]. In the presence of increased insulin resistance produced by over-weight and physical inactivity the reduction in insulin resistance around puberty may precipitate T2DM in these children especially if they have inadequate compensatory insulin secretion (see Figure 3.11).

Prenatal and early childhood events may also increase the risk of developing T2DM in children. Among siblings from the same nuclear family, the child of a mother who had diabetes during pregnancy has a threefold greater risk of developing early onset diabetes than their sibling who was born before the mother became diabetic [125]. This increased risk is mainly due to exposure to diabetes in utero since these siblings share the same environment and a similar probability of inheriting the same genetic composition.

Low birth weight and disproportionate growth in utero have been shown to be associated with both T2DM and insulin resistance [126,127]. Breast-feeding is also protective against obesity and T2DM [128–130] especially when sustained for a long durations.

Table 3.3 Challenges in the classification of diabetes in children

<table>
<thead>
<tr>
<th>Type 1 diabetes</th>
<th>Type 2 diabetes</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not overweight</td>
<td>Overweight or obese</td>
<td>Rising obesity means $&gt;30%$</td>
</tr>
<tr>
<td>Severe onset</td>
<td>Insidious</td>
<td>Presentations may vary</td>
</tr>
<tr>
<td>Weight loss</td>
<td>Some</td>
<td>Presentation may vary</td>
</tr>
<tr>
<td>Polyuria</td>
<td>Polyuria</td>
<td>Common to both</td>
</tr>
<tr>
<td>Polydipsia</td>
<td>Polydipsia</td>
<td>Common to both</td>
</tr>
<tr>
<td>Insulinopenia</td>
<td>No insulinopenia</td>
<td>Not present in all type 1 diabetes</td>
</tr>
<tr>
<td>Autoantibodies</td>
<td>No Autoantibodies</td>
<td>Usually occur in type 1, absence does not</td>
</tr>
<tr>
<td></td>
<td></td>
<td>exclude type 1</td>
</tr>
<tr>
<td>No insulin resistance</td>
<td>Insulin resistance</td>
<td>Usually occur in type 2</td>
</tr>
</tbody>
</table>
Complications and comorbid conditions

The development of T2DM in children is alarming. In adults, T2DM is associated with the development of complications such as retinopathy, renal disease, and CVD. Recent reports of T2DM in children suggest that diabetes carries a similar risk for the development of complications in early adulthood. In a study on the development of complications in Pima Indian children with T2DM, the rate of development of nephropathy was as severe in children as in adults over 30 years of follow-up [134]. The rate of development of retinopathy was, however, lower in cases that developed diabetes in childhood. A small series of young adults from the First Nation in Canada who had been diagnosed with T2DM before their 17th birthday were traced and followed up for the prevalence of diabetic complications and adverse outcomes by their 36th birthday. In this small series 9% had died, and 6.3% were already on dialysis and their blood glucose control was poor [132].

At diagnosis, a number of children have comorbid conditions such as dyslipidemia and hypertension both of which are risk factors for CVD [133]. Cardiovascular risk is strongly related to duration of diabetes [134] and T2DM in children will increase their risk to develop CVD later on in life.

The apparent rise in the prevalence of T2DM in childhood is presenting new challenges in the management and classification of diabetes in children. Adolescents are more likely to be noncompliant and may find it more difficult to follow the lifestyle modifications and treatment regimen that would be necessary to control blood glucose [108,135]. Furthermore, very few medications have been approved for use in treating type 2 diabetic children. Insulin and metformin are the only medications currently licensed for treatment of T2DM in children in the United States.

Modifiable risk factors for type 2 diabetes

Obesity

Obesity and weight gain have consistently been shown to be the one of the strongest modifiable risk factors for diabetes [140–142]. The ratio of a person’s weight in kilograms divided by the square of their height in meters called the body mass index (BMI) has been used in numerous studies as a surrogate for obesity. In a representative sample of the US population, each unit increase in BMI was associated with a 12% increased risk of T2DM [143]. Compared to people with BMI <22 kg/m² those with BMI of 25−27 kg/m² had 2.75 times the risk of diabetes, and each kilogram increase in body weight over 10 years was associated with a 4.5% increase in diabetes risk. Numerous studies have found similar results in different populations [144,145], but the magnitude of risk associated with a given BMI may differ across populations. The distribution of weight and weight gain are also important risk factor for diabetes. Central obesity, that is, deposition of fat in the trunk and abdominal areas, has been shown to be a strong risk factors for diabetes. The surrogate measures of central obesity include circumference of waist, waist-to-hip ratio, and waist-to-thigh ratio. More recent technological advances like dual-energy X-ray absorptiometry (DEXA) and magnetic resonance imaging (MRI), and computer tomography have made it possible to measure subcutaneous and intra-abdominal fat. Intra-abdominal fat has been shown to increase the risk of insulin resistance and diabetes in a number of studies in different populations [145–148] and this effect may be independent of the effects of total body obesity [149,150]. Obesity always occurs when energy intake exceeds energy expenditure.

Physical activity

Physical activity has been consistently reported to be inversely related to future risk of diabetes in most populations. Higher levels of physical activity are associated with a lower risk of diabetes in observational studies [151]. Increased physical activity reduces the risk of obesity but its effect on diabetes risk has been shown to be independent of its effect on body weight. Numerous studies have shown that exercise is related to acute and long-term improvements in insulin sensitivity and reduction in insulin concentrations [152,153]. Several epidemiologic studies in diverse populations [154–161] confirm a dose–response relationship between levels of physical activity and diabetes incidence. One study among Finnish men [160] found a strong gradient of diabetes risk associated with intensity of physical activity regardless of duration, but most evidence indicates that the total amount (i.e., number of days or minutes) of physical activity per week is a more important determinant.

Sedentary lifestyle

Sedentary lifestyle has also been shown to be a risk factor for both diabetes and obesity. Prolonged television (TV) watching as a surrogate of a sedentary lifestyle has also been reported to be a risk factor for diabetes [162,163]. Compared to men spending 0−1 h per week watching TV, those spending 2−10 h had 66% higher risk of diabetes, and the risk increased progressively to 187% higher incidence among those spending >40 h per week [162]. In women, every 2 h per day increment in TV watching was associated with a 23% increase in risk of
Epidemiology and geography of type 2 diabetes mellitus

obesity and 14% increase in risk of diabetes [163]. Furthermore, each 2-h increment in sitting at work was associated with a 5% increase in diabetes risk. It is estimated that a relatively active lifestyle (<10 h per week of TV watching and ≥30 min per day of brisk walking) can prevent 43% of new cases of diabetes [163].

Evidence from intervention studies
Results from the first diabetes prevention trials have clarified the role of physical activity as a risk factor for diabetes. Results from the Malmö trial in Sweden [164], Da Qing trial in China [58], Diabetes Prevention Study (DPS) in Finland [59], and the Diabetes Prevention Program in the United States [60] have all shown that moderate physical activity reduces the progression of IGT to diabetes by 30–58%. In these studies moderate physical activity was included in a lifestyle intervention arm of the trials with dietary changes that include reduction in caloric intake and in the percentage of dietary fat and increase in fiber intake (Table 3.4). In most of these studies it is difficult to discern the individual effect of the dietary intervention as it was combined with modifications in physical activity.

Dietary factors
Diet is a phenomenon that is made up of complex interactions between numerous foods and nutrients that are highly correlated and are determined by personal preference, cultural heritage, and socioeconomic factors. Diet is traditionally measured using questionnaires and food diaries. A number of studies have examined the relationship between diabetes and different aspects of diet including absolute intake of nutrients, nutrient intake as a percentage of total energy, dietary patterns, and bioavailability characteristics of foods like glycemic index.

Nutrients and diabetes risk
Dietary fat Animal studies and clinical human studies suggest several plausible mechanisms relating diet to the etiology of T2DM. High-fat diets have been associated with obesity [165], increased body fat for a given weight [166], and altered fat distribution [166]. Furthermore, alterations in cell membrane composition induced by the composition of dietary fat could alter membrane fluidity and/or insulin-mediated signal transduction as well as subsequent insulin action [167].

The results from epidemiologic studies on the relationship of total fat and T2DM have been controversial. In general, most of the migrant [168,169] and retrospective [170,171] studies show a positive relationship between high-fat, low-carbohydrate diets and T2DM, whereas results from prospective studies from different populations have been mixed [172–174] and not as consistent.

The effect of different subtypes of fat such as saturated, polyunsaturated, monounsaturated fat [175,176] and ω-3 fatty acids [177–179] on diabetes risk has also been investigated. An extensive review [180] concluded that while neither total fat nor total carbohydrate as proportions of total energy play a major part in the development of T2DM, specific types of fat and carbohydrate are important. For example, they concluded that (1) higher intake of polyunsaturated fat and long-chain n-3 fatty acids (fish oil) may be beneficial, and (2) higher intake of saturated fat and trans-fatty acids may be deleterious.

| Table 3.4 Lifestyle changes in the primary prevention of type 2 diabetes mellitus |
|-----------------|------------------|------------------|-----------------|-----------------|
| Study           | Population       | Age (years)      | Interventions                                           | Results                        | Comments                                |
| Malmö [164]     | 181 Swedish men with IGT | 47–49            | Diet, physical activity, and training                  | 50% with IGT reverted to normal; blood sugars normalized in 50% with diabetes | 6-year follow-up; participants were not randomized |
| Da Qing [58]    | 577 Chinese men and women with IGT | >25 (mean age 44.9) | Diet only, exercise only, and diet + exercise         | Diet only: 31% ↓ in diabetes risk; exercise only: 46% ↓; diet + exercise 42% ↓ | 6-year follow-up; randomization by clinic |
| Diabetes Prevention Study (DPS) [59] | 522 overweight Finnish men and women with IGT | Middle age (mean age 55) | Increased physical activity, increased fiber intake, and reduced fat intake | 58% decrease in diabetes risk | 3.2-year follow-up |
| DPP [60]        | 3324 overweight US men and women with IGT | Mean age 50.6    | Lifestyle (increased physical activity, weight loss, reduced fat intake), and drug (metformin) | Lifestyle 58% ↓ in diabetes risk; metformin 31% ↓ | 2.8-year follow-up; subjects from several different race/ethnic groups. Lifestyle interventions were effective in all race/ethnic and age groups |

Note: IGT = impaired glucose tolerance.
Dietary carbohydrate

The possibility of the risk of diabetes being related to carbohydrate consumption has been frequently raised. This hypothesis was generated because of the stronger, more direct, and more immediate challenge that ingestion of carbohydrate presents to the β cells, compared with ingestion of proteins or fats. Results from studies on carbohydrate and diabetes risk have, however, been inconclusive [169,181,182]. Other aspects of carbohydrate composition and metabolism including glycemic index and fiber content have also been examined.

Glycemic index The glycemic index [183] of food is a measure of the postprandial excursion of glucose as a result of the ingestion of a fixed amount of food. Foods with high glycemic index cause greater excursions of glucose. The postprandial excursions are dependent on the rate of absorption of glucose, which in turn is dependent on a number of factors including the type of carbohydrate (simple or complex) and the amount of fiber. Food with lower glycemic index has been associated with a lower risk of developing diabetes in some studies [182,184], but others have failed to show any association between the glycemic index of foods and the risk of developing diabetes [185,186].

Fiber intake High-fiber intakes have been related to a reduction in the risk of diabetes in some populations [182,184,187–190]. The mechanism by which high fiber reduces the risk of diabetes is unknown. High-fiber diets usually have a lower glycemic index but some studies that have shown a relationship between fiber intake and diabetes have failed to demonstrate a beneficial effect of the glycemic index of the diet [185]. Thiamine and vegetable protein are found in high concentrations in foods that are rich in fiber and have also been shown to reduce 2-h glucose concentration in women and may explain some of the association between fiber intake and diabetes [191,192].

Vegetable and fruit intake

Higher intakes of vegetables have been shown to be associated with a reduction in the risk of diabetes [172,173,193]. The component of vegetables that confers lower risk has not been clearly identified but may include antioxidants such as carotenoids and tocopherols [194], higher fiber intake, and vitamins [172,195,196]. The effect of fruit consumption on diabetes risk has been inconclusive [173,193].

Dietary patterns and scores

Studies have examined the effect of the whole composition and quality of diet on diabetes risk using various scores and dietary patterns identified by principal component and factor analyses [197,198] on diabetes risk. The results from these studies have shown that a diet that has relatively higher intakes of vegetables and fruits and lower intakes of fat-rich foods has been associated with a reduction in diabetes risk [197,198].

Self-rated dietary patterns can be used to discriminate differences in the nutrient composition in diet. Pima Indians who self-reported a diet more similar to their traditional Pima diet had a lower risk of developing diabetes over a 12-year period [199] when compared to a more Westernized dietary preference. The traditional Pima diet had a higher fiber and lower fat intake when compared to the more Westernized diet. The traditional diets of most indigenous populations have a higher intake of fiber and lower glycemic index [200,201] than the more recent Westernized diets that most of these populations have adopted.

Nontraditional modifiable risk factors for type 2 diabetes

Inflammation

Although insulin resistance and relative insulin deficiency represent the main characteristics of T2DM, the underlying mechanism responsible for these abnormalities remains largely unknown. On the basis of hypothesis that both T2DM and atherosclerosis may generate from a “common soil” [202,203], more recently a growing body of evidence points out that inflammation may constitute the common factor that leads to the development of both these diseases. One of the most used markers of subclinical inflammation is the C-reactive protein (CRP). A number of cross-sectional studies have shown that increased concentrations of CRP are associated with abnormalities characterizing the metabolic syndrome, including obesity, insulin resistance, low HDL cholesterol, and hypertriglyceridemia [204–206]. In addition, prospective studies have demonstrated that high levels of CRP increase the risk of developing T2DM [207–212]. These findings have opened new avenues for understanding the pathogenesis of T2DM and, eventually, for preventing the disease. Indeed, if, as these data have indicated, subclinical inflammation is an important determinant of T2DM, then the use of anti-inflammatory drugs may prevent diabetes. In this regard, it is noteworthy that in patients with T2DM high-dose aspirin reduces insulin resistance and improves glucose tolerance [213].

Smoking

Numerous prospective studies have indicated that smoking is associated with the development of diabetes. In the US Nurses’ Health Study, women who smoked at least 25 cigarettes per day compared to those who never smoked had a relative risk of developing diabetes of 1.42, even after controlling for known risk factors [214]. Similar results were shown in men [215]. These findings were confirmed by Will et al. in a prospective study involving over 275,000 men and 434,000 women from the US Cancer Prevention Study [216]. Among those who smoked at least two packs per day at baseline, men had a 45% higher diabetes rate than men who had never smoked; in women the comparable increase was 74%. More important, quitting smoking reduced the rate of diabetes to that of nonsmokers after 5 years in women and after 10 years in men. In support of these findings are cross-sectional data on the association between cigarette smoking and hemoglobin A1C (HbA1C)
from the East Anglian component of the European Prospective Investigation into Cancer (EPIC-Norfolk) [196]. In this study, current smokers had highest mean HbAIC concentrations, lowest levels were observed in never smokers, and intermediate in former smokers. HbAIC levels also correlated in a dose–response manner both with the number of cigarettes smoked per day and with total amount of smoking as measured by pack-years. This association persisted even after adjusting for potential confounders including BMI, waist-to-hip ratio, physical activity, and dietary variables. However, to better assess the role of smoking as a determinant of T2DM we need cohort studies to further confirm that this association exists in different populations.

There is evidence that smoking is associated with insulin resistance [217,218]. In a study by Facchini et al. [217] chronic smokers compared to nonsmokers had significantly higher plasma triglycerides and lower HDL-cholesterol levels and higher insulin concentrations after a 75-g oral glucose challenge. Well-designed clinical studies of the effects of acute and chronic smoking on insulin resistance are needed to elucidate the mechanism by which smoking induces insulin resistance. It is plausible that oxidative stress caused by smoking may induce endothelial dysfunction, resulting in insulin resistance into muscle and liver.

The possibility that smoking may play a causal role in the development of T2DM has important implications for prevention. Both diabetes and smoking are conditions common enough that even a small increase in the risk of diabetes associated with smoking may have an important public health impact.

References


95 Owada M, Hanaoka Y, Tanimoto Y, Kitagawa T: Descriptive epidemiology of non-insulin dependent diabetes mellitus detected by


105 Hales CN, Barker DJ, Clark PM, et al.: Fetal and infant growth and impaired glucose tolerance at age 64. *Arch Dis Child* 2000;83:373–379.


204 Yudkin JS, Stehouwer CD, Emeis JJ, Coppack SW: C-reactive protein in healthy subjects: associations with obesity, insulin resistance, and endothelial dysfunction: a potential role for


SECTION II

Physiology of pancreatic function
CHAPTER 4

Development and maintenance of the islet β cell

Teresa L. Mastracci, Raghavendra G. Mirmira and Ryan M. Anderson
Department of Pediatrics, Indiana University School of Medicine, Indianapolis, IN, USA

Key points

• The islets of Langerhans are composed of five hormone-secreting cell types: α cells produce glucagon, β cells produce insulin, δ cells produce somatostatin, PP cells produce pancreatic polypeptide, and ε cells produce ghrelin.
• The pancreas is derived from the merger of distinct buds that are induced dorsally and ventrally from the foregut endoderm by the adjacent mesoderm.
• The extensively branched structure of the mature pancreas is attained by continued growth and remodeling of the primitive tubular epithelium in both dorsal and ventral buds.
• The branching epithelium becomes organized into “tip” and “trunk” domains, within which progenitor cells reside.
• The “secondary transition” is the phase during mid-pancreas development that is marked by the differentiation of exocrine cells and the major wave of islet cells.
• A number of transcription factors expressed broadly in early pancreas development later become restricted to specific endocrine cell types and subsequently acquire additional functions in these cell types.
• The state of β-cell mass and function is a large determinant of glucose and lipid homeostasis.
• β cells exhibit progressively slower rates of turnover as a mammal ages.

Introduction

The discovery of insulin by Banting and Best in the early 1920s marked the beginning of a new era of diabetes research that focused on the study of insulin and the biology of the hormone-producing cells of the pancreas. In the ensuing decades, although much was learned regarding the synthesis, structure, and action of insulin, the developmental origins of insulin-producing β cells remained ambiguous. In the 1970s, the long-held belief that β cells might arise from the neural crest was refuted by elegant interspecies cell transplantation studies, which demonstrated that radiolabeled quail neural crest cells do not populate pancreatic endocrine tissues in host chick embryos [1]. This seminal finding, and the advent of techniques to manipulate the mouse genome, has resulted in a greater understanding of how the pancreas develops and matures. The lessons from developmental biology have also elicited speculation that maintenance of endocrine cell mass in the adult pancreas follows a similar paradigm, whereby a pool of proliferative progenitor cells differentiates as needed to repopulate the endocrine compartment. It is becoming increasingly recognized that processes that maintain or increase β-cell mass rely on signals distinct from those in the developing pancreas, and therefore the mechanisms of cell specification and maintenance differ depending upon the age of the organism. This chapter reviews the fundamentals of pancreas organogenesis, proceeds with an overview of cytodifferentiation with an emphasis on the β cell, and concludes with a review of β-cell growth and maintenance in the adult pancreas.

Pancreas development

The pancreas is a compound digestive gland, comprised of both exocrine and endocrine components that secrete digestive enzymes and hormones, respectively. The exocrine component makes up approximately 95% of pancreas mass, and consists of acinar cells and duct cells. Acinar cells produce and secrete proteases, lipases, amylases, and nucleases that are necessary for the digestion of nutrients. Duct cells form a tubular network throughout the pancreas, and also secrete mucins and fluids that flush the acinar secretions to the intestine. The mature endocrine component of the pancreas, organized into structures known as the islets of Langerhans, comprises ~2%
of the total organ mass. In the adult, islets of Langerhans
are composed primarily of five discrete hormone secreting
cell types: α cells produce glucagon, β cells produce insulin
and islet amyloid polypeptide, δ cells produce somatostatin,
PP cells produce pancreatic polypeptide, and ε cells produce
grelin. β Cells comprise the majority (60–80%) of the cells
that make up the islet in vertebrate animals. The progression
of pancreas development in vertebrates can be segmented into
five major events (summarized in Figure 4.1): (1) induction of
definitive endoderm, (2) formation of the primitive gut tube
and patterning of endoderm into organ-specific progenitor
zones, (3) induction of dorsal and ventral pancreatic buds, (4)
outgrowth, branching, and fusion of the pancreatic buds, and
(5) cytodifferentiation.

**Induction of the definitive endoderm**

During early development, pluripotent stem cells in the gastrula
stage embryo evolve into the multipotent progenitor cells of the
three primary germ layers known as ectoderm, mesoderm, and
definitive endoderm. Broadly speaking, mesodermal derivatives
provide support and movement, whereas ectodermal derivatives
provide for sensation of, and protection from, the environment.
The definitive endoderm is the innermost germ layer of meta-
zooan embryos, and its derivatives mediate exchanges with the
environment, including nutrient absorption and gas exchange.
Definitive endoderm gives rise to the pancreas, liver, intestine
and other digestive organs, as well as the thyroid and parathyroid
glands and the respiratory tract.

Model systems must be employed to study the mechanisms
of development. Considering its evolutionary proximity
to humans and the wealth of sophisticated tools for genetic
manipulation, the mouse (*Mus musculus*) is the most pervasively
utilized model organism. However, much of what is known
about endoderm induction and morphogenesis has been
learned through the study of lower vertebrates. Model systems
such as the zebrafish (*Danio rerio*), the African clawed frog
(*Xenopus laevis*), and the chicken (*Gallus gallus*) each exhibit
species-specific experimental advantages (see Table 4.1).

During gastrulation, cells delaminate from the epiblast cell
layer, and ingress through a transitory structure called the
primitive streak in amniotes (e.g. mammals and birds) or the
marginal zone in fish and frog, and emerge as endoderm cells.
Although the morphogenetic events of gastrulation vary widely
in model organisms, these events are regulated by a conserved
set of regulatory molecules. The most fundamental endoderm
induction signal is Nodal, a secreted TGFβ-like molecule that
is highly conserved in amniotes, amphibia, and fish. Nodal
signaling is necessary and sufficient in all vertebrates to trigger
a genetic cascade leading to endoderm formation [2]. Nodal is
expressed at the site of gastrulation, where it initially generates
a transitory mesendodermal precursor cell population. This
population is subsequently subdivided into mesoderm and
endoderm by the morphogenetic properties of Nodal, with the
highest levels of Nodal signaling producing endoderm.
Expression of the gene encoding Nodal is initiated in diver-
gent ways depending upon the species. In mice, the gene is
induced by WNT signaling at the interface of embryonic and
extra-embryonic tissues [3], whereas in zebrafish and frogs it is
induced by maternally provided factors that are localized to the
site of Nodal induction [4,5]. However, in all vertebrates studied,
Nodal expression is enhanced by high levels of WNT signaling and maintained by a positive paracrine feedback loop [6]. The targets of Nodal include the Mix-like (Mixl) family of homeobox genes, the Foxa family of forkhead genes, and the high mobility group box gene Sox17. Together, these genes comprise an essential transcription factor network that stabilizes the endodermal fate while simultaneously segregating it from mesoderm [7,8].

**Patterning of the gut tube into organ domains**

All endoderm generated during gastrulation is not equivalent: cells acquire positional identity upon specification, which is dependent upon the time at which the cells are specified. The foregut endoderm is specified first, followed by midgut, then hindgut. As gastrulation concludes, the foregut expresses unique spatial markers, including Hhex, which is directly induced by Nodal, as well as Sox2 and Foxa2. These three transcription factors are essential for the development of foregut organs [2,9,10]. Moreover, the antero-posterior (A-P) patterning of the endoderm is dictated by the release of several signaling molecules from the mesoderm. In addition to Nodal, which induces anterior endoderm at the highest levels of expression and posterior fates at progressively lower levels, a core set of secreted signaling molecules are encountered by migrating cells. Fibroblast growth factors (FGFs), Wingless/Int (WNTs), bone
morphogenetic proteins (BMPs), and retinoic acid (RA) are secreted from a posterior location, serve to suppress anterior fates, and to partition further the gut tube into regions that express genes encoding the crucial transcription factors Pdx1 (intermediate gut region) and Cdx1/2/4 (hindgut region) [3,8]. Moreover, as discussed later, this set of signaling molecules acts iteratively within the pancreatic endoderm throughout development to generate the mature pancreas.

Immediately after formation, the germ layers are arranged as three flat, stacked sheets of cells. The endoderm rapidly transitions to a tubular morphology and becomes enveloped by mesoderm. In amniotes, the morphological transformation begins with the folding together of the lateral edges of the anterior and posterior ends of the sheet to form foregut and hindgut pockets. The lateral edges of the endoderm sheet are then fused together by zipper-like morphogenetic behavior at the ventral midline that progressively seals from the anterior and posterior ends; this process ultimately generates a hollow gut tube. In zebrafish and frogs, the endodermal sheet first forms a solid rod through mass endodermal cell migration toward the midline. This rod later hollows by the mechanism of cavitation, which may involve lumen-directed fluid transport.

The foregut endoderm gives rise to multiple organs, including pancreas, liver, thyroid, and lung. The gut tube becomes subdivided into organ forming regions through reciprocal interactions between the endodermal cells and the adjacent mesenchyme, a process requiring FGF, BMP and RA signaling pathways. Explant studies using uncommitted mouse foregut originally revealed the role of FGF emanating from the septum transversum mesenchyme. Low levels of FGF led to expression of Pdx1 in ventral pancreatic progenitors, intermediate levels led to hepatic fates, whereas the highest levels specified lung and thyroid. Mouse genetic experiments later confirmed this model in vivo, and the requirement for FGF is conserved in both zebrafish and frogs [4,5,11]. BMP signals, which also originate from the septum transversum mesenchyme in mammals and from the lateral plate mesoderm in fish, regulate pancreatic specification [6,12]. BMP signals repress pancreatic development from a common pool of progenitors [7,8,13]. RA signaling biases endoderm to more posterior fates, and may have a role in partitioning the expression domains of the genes encoding Pdx1 (Pancreatic and duodenum homeobox 1) and Cdx1 (Caudal type homeobox 1), which are considered master regulators of pancreatic and intestinal fates, respectively. Furthermore, RA may have a direct role in dorsal bud induction, as both mice and zebrafish deficient in RA signaling lack the dorsal pancreas [14,15].

**Induction of the dorsal and ventral pancreatic buds**

The pancreas is derived from the merger of distinct buds that are induced dorsally and ventrally from the foregut endoderm by the adjacent mesoderm. In all vertebrates, both dorsal and ventral buds are marked by expression of Pdx1 and Ptf1a (Pancreatic transcription factor 1a), two transcription factors essential for proper development of all pancreatic cell types [16]. The ventral bud is specified by low-level FGF signaling from the mesenchyme in the absence of BMP signaling, as described earlier. Dorsal bud fate is specified when FGF2 and Activin signals secreted from the notochord mitigate Sonic Hedgehog (Shh) expression in the dorsal midline of the gut tube [17]. In contrast to the suppressive role of Shh in early pancreatic bud induction, Shh later plays a positive role in the expansion of the pancreatic epithelium and in β-cell function [18]. Additionally, dorsal bud induction requires RA [19] as well as yet unidentified factors secreted from the dorsal aorta. In contrast to amniotes and frogs, dorsal fate specification likely occurs by a divergent mechanism in zebrafish, as Shh has a positive regulatory role in dorsal bud induction [20]. Furthermore, signaling from the dorsal aorta does not appear to have a fundamental role in zebrafish, as cloche mutants of zebrafish that lack all vasculature still demonstrate normal bud induction [21].

**Pancreatic growth and epithelial branching morphogenesis**

The extensively branched structure of the mature pancreas is attained by continued growth and remodeling of the primitive tubular epithelium in both dorsal and ventral buds. Once the buds have been induced, they are critically dependent on interaction with the mesenchyme for continued morphogenesis and cytodifferentiation. Moreover, during tubulogenesis, cytodifferentiation is suppressed, which allows for the expansion of progenitor cell populations, and implies direct communication between the genetic programs controlling each process. Mouse explant studies showed that isolated pancreatic epithelium failed to develop acinar tissues in the absence of mesenchyme, and that secreted factors were likely the mediators of this development [22]. In complementary studies in vivo, targeted ablation of pancreatic mesenchyme showed essential roles for epithelial-mesenchymal signaling during both early and late bud morphogenesis [23]. FGF10 from pancreatic mesenchyme supports the proliferative expansion of the epithelium as well as the maintenance of undifferentiated pancreatic progenitor cells via induction of the Notch pathway [24,25]. Additional signals from the mesenchyme instruct the continued development of the buds. For instance, reciprocal EphB signaling between mesenchyme and epithelium is positively required for branching, growth, and cytodifferentiation [26], while unidentified signals from blood vessels restrain these processes [27]. Even as these and other unknown signals regulate its growth, pancreatic mass is ultimately constrained by an intrinsic program established in the pancreatic progenitor domain [28]. Finally, as the foregut grows and elongates, the developing ventral pancreatic tissues rotate along with the gut, and ultimately fuse with the dorsal bud to generate mature pancreatic architecture. Although congenital malformations arising from defective pancreatic bud fusion are relatively common, including annular pancreas and...
pancreas divisum, the genetic pathways underlying this process are not well understood [29].

**Cytodifferentiation in the developing pancreas**

As stated earlier, the early evaginating pancreatic buds are made up of progenitor cells that express Pdx1 and Ptf1a. These early pancreatic progenitor cells will induce the specification of multipotent progenitor cells (MPCs), which direct derivative cell populations toward a particular fate. The early developing pancreatic buds are marked by the appearance of cells with low-level digestive enzyme production and an initial wave of glucagon- and insulin-expressing cell types, a period referred to as the “primary transition” of pancreas formation. The term “secondary transition” is applied to the phase during mid-pancreas development that is marked by the differentiation of exocrine cells and the major wave of islet cell formation [30]. The secondary transition is characterized by a dramatic increase in cells expressing acinar digestive enzymes, as well as a large increase in cells producing endocrine hormones including insulin, glucagon, ghrelin, somatostatin, or pancreatic polypeptide. Preceding this major wave of differentiation, the secondary transition also encompasses the emergence of the MPCs and the establishment of the pre-acinar and bipotent duct/endocrine cell populations from which the differentiated exocrine, and endocrine or duct cells derive, respectively (Figure 4.2).

In recent years, a network of genes has been identified whose products specify the development of the different cell types (see Table 4.2). The importance of these genes has largely been identified by lineage tracing studies and targeted mutations in mouse models, and has led to two important concepts in pancreatic cytodifferentiation. First, the products of many of these genes function in a “cell-autonomous” manner, meaning that their expression level in a given cell type alters the fate and function of that cell. Second, misexpression of specific genes in magnitude or in a spatial (cell- or domain-specific) or temporal (time of development-specific) manner can redirect developing progenitor cells to cell fates they would otherwise not have adopted. With respect to the latter concept, the study of cell-autonomous factors has the potential to identify means through which other cell types might be converted to β cells for the treatment of different forms of diabetes.

**Tip versus trunk domains**

The stratification of the pancreatic epithelium and the resulting formation of microlumens is an essential morphological change that occurs during the primary transition. The subsequent remodeling of the ductal plexus and branching of the epithelium continues throughout embryonic pancreas development [26]. Epithelial branching leads to the morphogenesis of different domains, which become most apparent at the beginning of the secondary transition. There is mounting evidence that MPCs exist within these emerging epithelial domains. In particular, Melton and colleagues proposed the identity of the MPCs as those cells expressing the factors Ptf1a, Pdx1, c-Myc, and Cpa1 [31]. This concept stemmed from a genome-wide transcription factor analysis in mouse pancreas tissue at embryonic day (E) 14.5, whereby gene expression patterns were identified to segregate into particular domains of the developing pancreas. Specifically, patterns emerged
Table 4.2 Key transcription factors in β-cell development

<table>
<thead>
<tr>
<th>Alternate names</th>
<th>Mouse knockout pancreas/β-cell phenotypes</th>
<th>Human digestive disease relevance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hhex</td>
<td>loss of foregut endoderm, specification of ventral pancreas</td>
<td>none associated</td>
</tr>
<tr>
<td>Foxa2</td>
<td>loss of foregut endoderm, insulin hypersecretion</td>
<td>type 2 diabetes susceptibility</td>
</tr>
<tr>
<td>Sox17</td>
<td>depletion of definitive endoderm</td>
<td>none associated</td>
</tr>
<tr>
<td>HNF1β</td>
<td>required for visceral endoderm and gut morphogenesis</td>
<td>MODY5</td>
</tr>
<tr>
<td>Pdx1/Ipf</td>
<td>pancreatic hypoplasia, β-cell secretory dysfunction</td>
<td>MODY4/type 2 diabetes susceptibility, pancreatic hypoplasia</td>
</tr>
<tr>
<td>Ptf1a</td>
<td>pancreatic hypoplasia and transdifferentiation of pancreas to duodenum</td>
<td>pancreatic hypoplasia</td>
</tr>
<tr>
<td>Mnx1</td>
<td>dorsal bud agenesis, reduced β-cell mass</td>
<td>neonatal diabetes</td>
</tr>
<tr>
<td>Sox9</td>
<td>pancreatic hypoplasia, early differentiation</td>
<td>none associated</td>
</tr>
<tr>
<td>Neurog3</td>
<td>lack of all endocrine cell types</td>
<td>congenital malabsorptive diarrhea with variable onset diabetes</td>
</tr>
<tr>
<td>Neurod1</td>
<td>diminished endocrine cell mass and diabetes</td>
<td>MODY6</td>
</tr>
<tr>
<td>Isl1</td>
<td>dorsal bud agenesis, lack of islets</td>
<td>neonatal diabetes</td>
</tr>
<tr>
<td>Pax6</td>
<td>reduction of β cells</td>
<td>none associated</td>
</tr>
<tr>
<td>Arx</td>
<td>loss of α cells and increase of β cells</td>
<td>none associated</td>
</tr>
<tr>
<td>Pax4</td>
<td>loss of β cells and gain of α cells</td>
<td>MODY9</td>
</tr>
<tr>
<td>Nkx2.2</td>
<td>arrested β-cell differentiation</td>
<td>none associated</td>
</tr>
<tr>
<td>Nkx6.1</td>
<td>loss of β cells in the secondary transition</td>
<td>none associated</td>
</tr>
<tr>
<td>Mafa</td>
<td>normal differentiation of β cells, impaired function</td>
<td>none associated</td>
</tr>
<tr>
<td>Mafb</td>
<td>embryonic lethal; reduced numbers of β and α cells</td>
<td>none associated</td>
</tr>
</tbody>
</table>

that could be grouped into five domains: pan-epithelium, tip, trunk, mesenchyme, and vasculature. Genes discovered to be expressed in the tip domain later segregated into differentiated acinar cells, whereas genes expressed in the trunk domain were identified in the ducts or differentiated endocrine cells. Taken together, data from multiple studies suggest that MPCs residing in the tip domain will give rise to pre-acinar cells, destined to become exocrine tissue, and bipotent duct/endocrine cells that reside in the trunk of the branching epithelium (Figure 4.2).

**Endocrine versus exocrine cell fate decision**

It is in the early developing pancreatic domain, when progenitor cells are multipotent, that the endocrine versus exocrine decision is made. In the mouse, Ptf1a is located in the early pancreatic progenitor cells and over time becomes restricted to expression in the branching tips and then differentiated acinar cells [32]. At the beginning of pancreas development the transcription factors Nkx6.1 and Nkx6.2 are co-expressed in the MPCs before becoming restricted and separated in their expression pattern. The Nkx6 factors and Ptf1a have been noted to function antagonistically in the decision between endocrine and exocrine cell fates, such that Nkx6 factors promote the endocrine decision whereas Ptf1a promotes the exocrine decision [33].

The endocrine versus exocrine cell fate decision is also influenced by the level of expression of the transcription factor Neurogenin3 (Neurog3) in the progenitor cells. Specifically, a high level of Neurog3 is required for commitment to the endocrine fate [34]. Moreover, Notch signaling is used in the trunk domain to subdivide this compartment between endocrine and ductal cells via a lateral inhibition mechanism. Neurog3 upregulates expression of the Notch ligand Delta-like 1 (Dll1) in endocrine progenitors, which activates the Notch pathway in neighboring cells thereby repressing their differentiation into endocrine cells.

**The endocrine progenitor cell**

The culmination of many studies has confirmed that in the developing mouse pancreas, the transcription factor that defines the endocrine progenitors is Neurog3. Neurog3-null mice exhibit absence of endocrine cells in the pancreas, and such mice develop neonatal diabetes and die shortly after birth [35]. During mouse pancreas development a subset of hormone-expressing cells is observed as early as E9.5, whereas the major wave of endocrine differentiation occurs during the secondary transition. Lineage tracing experiments using genetically-engineered mouse reporter lines identified that, regardless of when the endocrine cell differentiates, all hormone-expressing cells are derived from cells that previously expressed Neurog3 [36,37].

The process of endocrine differentiation has also been linked to the morphological process of delamination of the progenitor cells from the pancreatic epithelium. Interestingly, the delamination of progenitor cells is initiated in the cells that express
Neurog3 [38]. Moreover, the subsequent differentiation into different endocrine cells types is influenced by the timing of Neurog3 expression. Specifically, altering the temporal expression of the gene encoding Neurog3 in the mouse influences the competence of progenitor cells to differentiate into the specific endocrine cell types, such that earlier expression produces almost exclusively α cells, whereas later expression produces varied ratios of all hormone-expressing cell types [39].

Previous models of pancreas development suggested that each Neurog3-expressing cell could give rise to any subsequent differentiated endocrine cell type. However, this perspective has been challenged by lineage tracing experiments using genetically-altered mice, which demonstrated that each Neurog3-expressing endocrine progenitor cell is in fact unipotent, and therefore destined to become a particular single-hormone expressing endocrine cell type [40] (Figure 4.2). The implication of this discovery is that the transcription factor “code” responsible for the differentiation of each hormone-expressing cell type may be delineated before endocrine progenitors are specified.

Clearly, the expression of Neurog3 is of great significance to the development and differentiation of endocrine cells in the mouse. However, the effect of loss of this transcription factor in other species is not identical to the mouse. For example, in zebrafish Neurog3 is not observed in the pancreas [41]. Homozygous mutations in Neurog3 have been identified in humans, resulting in congenital malabsorptive diarrhea and childhood-onset diabetes [42,43], but without congenital loss of pancreatic endocrine cells (as seen in the mouse). Nevertheless, the absence of enteroendocrine cells was noted in these individuals.

Other transcription factors are also expressed in the early endocrine cell population, and genetic deletion studies identified these factors to be crucial to endocrine cell differentiation. In particular, the transcription factor Islet1 (Isl1) is expressed in all mature, non-replicating islet cell types. Interestingly, Isl1 expression is also observed in the mesenchyme that surrounds the early dorsal pancreatic bud. The dorsal pancreatic mesenchyme does not form in Isl1-deficient mouse embryos, leading to a loss of extrocrine differentiation in the dorsal pancreas; the pancreas is also devoid of all islets in these mice. Loss of the transcription factor Pax6 in the mouse leads to death shortly after birth. The pancreas of Pax6-deficient animals is devoid of α cells and has marked reductions in β, δ, and PP cells [44]. Whereas a human mutation in the ISL1 gene has been identified in a patient with type 2 diabetes [45], no link to diabetes has been observed in humans with mutations in the PAX6 gene. Therefore similar to Neurog3, the functional importance between lower organisms and humans may not be completely conserved for genes involved in endocrine differentiation.

What makes a β cell?
The insulin-producing β cell is perhaps the most intensely studied endocrine cell type, largely because of the implications for understanding the pathogenesis and treatment of diabetes. Many mouse models have clarified the factors necessary for β-cell differentiation, development, and maturation (see Table 4.2). One such factor is the basic helix-loop-helix transcription factor Neurod1. Interestingly, Neurod1 is expressed in all endocrine cell types except the somatostatin-producing δ cell, and targeted disruption of the Neurod1 gene in mice results in severe reduction in α and β cells, and in neonatal diabetes owing to β-cell apoptosis [46]. The Maf family of transcription factors is also involved in the pathway of α- and β-cell differentiation. Both MafA- and MafB-deficient mouse models have pancreatic phenotypes. Loss of MafB leads to perinatal lethality and, although the total endocrine cell mass is unaffected, the pancreas shows reduced numbers of α and β cells [47]. By contrast, mice with a targeted deletion of gene encoding MafA are born viable and with normal islet cell numbers, but demonstrate β-cell dysfunction with advancing age, leading to glucose intolerance and diabetes [48]. In these mice, β-cell genes, including those encoding insulin, Neurod1, and Glut2, are significantly reduced. Owing to its importance in β-cell function, MafA is considered as a marker for mature, functional β cells.

Interestingly, a number of factors expressed broadly in early pancreas development become restricted to specific endocrine cell types during the secondary transition and acquire additional function in the differentiation or maturation of these cell types. One particular example is Pdx1. Whereas mice with a homozygous deletion of the gene encoding Pdx1 are born without a pancreas, haploinsufficiency of Pdx1 results in glucose intolerance [49,50]. Virtually identical phenotypes are observed in humans with respective homozygous and heterozygous mutations in Pdx1 [44]. After the pancreas is fully developed, Pdx1 is expressed primarily in β cells and is necessary for β-cell function, including transcriptional activation of several β-cell genes [51]. Nkx2.2 is another transcription factor that shows expression in the early pancreatic progenitors but becomes restricted to specific endocrine cell populations later in development [52]. Loss of Nkx2.2 in the mouse results in the complete absence of differentiated insulin-producing β cells, a significant decrease in α and PP cells, and a concomitant increase in ghrelin-expressing ε cells [53].

Mouse models have also identified specific factors necessary for the differentiation of the glucagon-expressing α cell. In particular, deletion of the transcription factor Arx results in hypoglycemia and neonatal lethality. Given that Arx is expressed in all endocrine cells except the β cell, the pancreas of the Arx-null mouse has altered endocrine cell ratios: a complete absence of α cells and an increase of β and δ cells. Conversely, the mis-expression of the Arx gene in either Pdx1- or Pax6-expressing cells results in a loss of β and δ cells and an increase in α and PP cells [48]. Moreover, compound mutants have demonstrated the complexity of transcription factor interactions and the importance of Arx function to endocrine cell development. Specifically, deletion of the Arx and Pax4 genes result in the loss of α and β cells but an increase in δ cells [48], and the Arx/Nkx2.2
compound mutant pancreas showed a restoration of the PP cell population that was lost in the Nkx2.2 null pancreas [52].

The evidence for the relevance of transcription factors to human pancreas development and disease is highlighted by the discovery that mutations in a number of transcription factors identified to be important in pancreas development in the mouse also display pancreas-related phenotypes in the human. Mutations or deletions of many of these factors have been established as the cause of monogenic forms of diabetes known as maturity-onset diabetes of the young (MODY) [54], or as the cause of human syndromes that include diabetes [42,43,55–58] (Table 4.2).

Translating pancreas development into cell-based therapies for diabetes

The knowledge gained from decades of pancreas development research has stimulated the translational pursuit of engineering insulin-producing β cells in vitro for therapeutic purposes. Specific extrinsic factors, including FGF, RA, and inhibitors of BMP or Shh signaling, have been applied to mouse and human embryonic stem cells in culture to successfully drive these malleable cells toward a pancreas fate [54] (see Figure 4.3). More recently, success in the creation of pluripotent, embryonic-like stem cells from somatic cells has opened the possibility of generating patient-specific β cells as cell replacement therapies for diabetes [55]. To date, however, such techniques have resulted in compound or mixed populations of hormone-producing cells, which have little or no capacity for glucose-stimulated insulin secretion when generated wholly in vitro, suggesting that many intrinsic and/or extrinsic factors involved in the cytodifferentiation of pancreatic progenitors remain to be identified.

In the mouse, the genetic manipulation of certain intrinsic factors, that is, Pdx1, Neurog3, MafA [59], or the induction of severe pancreatic injury [60], has also identified the capacity of differentiated cells in the pancreas to be “reprogrammed” to β cells. Therefore the continued merging of developmental biology research with in vitro differentiation technology may produce the long awaited therapeutic cure for diabetes.

Postnatal β-cell growth and maintenance

Following the secondary transition, the total mass of the pancreas increases substantially, but β cells comprise only about 1–2% of this cellular mass in the mature adult pancreas. Despite this relatively small percentage, the states of β-cell mass and function represent perhaps the greatest determinants in overall glucose and lipid homeostasis in virtually all types of metabolic disorders [61]. Two fundamental concepts regarding postnatal β cells have emerged over the past two decades: (1) although β cells were considered to be a postmitotic cell type in an adult mammal, it is now understood that they indeed exhibit a slow rate of turnover that decreases with age [62], and (2) the mass and function of β cells (and therefore the balance between new cell formation and death) can be dynamically altered to an extent to compensate for the physiologic or pathologic state of the organism [60]. A major focus and area of controversy has been the mechanisms that underlie postnatal β-cell growth and maintenance. From the discussion in the preceding sections on prenatal development, it is evident that formation of most β cells during embryogenesis occurs through a process known as neogenesis, in which new cells arise from the differentiation of stem or progenitor cells. Although some studies suggest the existence of multipotent stem cells within the postnatal rodent pancreatic epithelium, such cells normally do not give rise to significant numbers of new β cells in adult animals.

Neonatal β-cell turnover

The neonatal period between birth and weaning in rodents is characterized by a high rate of β-cell turnover and net increase in β-cell mass. Turnover is defined as the dynamic formation and loss of cellular mass [62]. Because there is no

![Figure 4.3](image-url)  
**Figure 4.3** *In vitro* differentiation of pancreatic β cells. A schematic representation of a five-step embryonic stem (ES) cell differentiation protocol that mimics *in vivo* events of pancreatic organogenesis and differentiation through addition of secreted signaling factors. Important tissue markers that are shared with development of embryonic tissues are listed below each stage. (1) ES are guided through the mesendoderm (ME) stage to definitive endoderm (DE), then (2) to primitive foregut tube (GT), (3) posterior foregut (PF), and (4) pancreatic endoderm (PE). (5) The factors regulating the final stage of differentiation into endocrine cells (EN), cannot be performed effectively *in vitro*, and cells are transplanted into a murine host to complete cytodifferentiation by receiving physiologic cues. Additional abbreviations: FBS, fetal bovine serum; Cyc, cycloamine (SHH antagonist); Nog, Noggin (BMP antagonist); B27, defined tissue culture supplement. Source: Kroon E, et al. *Nature Biotechnology* 2008;26:443–452. Reproduced with permission of Nature Publishing Group.
longitudinal, noninvasive way to measure β-cell turnover in a given animal, the techniques that estimate β-cell turnover are based on cross-sectional studies from cohorts of animals that analyze steady-state β-cell mass, new β-cell formation (primarily by replication), and β-cell death. Nonetheless, studies using thymidine analog (BrdU) incorporation estimate that in the neonatal rat the rate of β-cell replication is as high as ~20% new cells per day at 2 days of age, and falling to about ~10% new cells per day by the time of weaning [63]. By contrast, replication rates in adult rats and mice are much lower, in the range of 0–2% new cells per day [63,64]. Although replication is thought to be the primary source of new cells per day by the time of weaning [63]. By contrast, the neonatal rat the rate of β-cell formation during this period, studies of thymidine analog incorporation cannot detect specific contributions from neogenesis, which is thought to play a role during the neonatal period [65]. Balancing this rate of replication is, in part, the rate of apoptosis that appears to be elevated during the neonatal period, with the frequency of apoptotic cells rising as high as ~4% (compared to less than 0.4% in adult rats) [63]. However, it should be noted that the true rate of β-cell death is very difficult to measure because dead/dying cells may be cleared more rapidly than can be measured by tissue morphometry, and other forms of death (necrosis) are not typically measured.

The foregoing studies in rodents appear to also reflect the dynamics of β-cell turnover in humans. Based on autopsy studies, the replication rate of β cells appears highest in children, especially infants (coincident with increases in β-cell mass during early life), then declines in adulthood [66]. Taken together, these studies suggest a dynamic remodeling of β cells and their mass in the neonatal/early postnatal period, and mechanisms underlying the increase in β-cell replication rate have been the focus of intense investigation [65]. The growth factors insulin and insulin-like growth factors (IGFs) are obvious candidates, given the autonomous production of both by β cells in the early postnatal period. Elimination of IRS-2 (a key protein in growth factor signaling) results in the failure to maintain β-cell mass in the face of increasing insulin resistance as mice age, suggesting a potential β-cell growth-promoting effect of this signaling cascade [67]. However, elimination of either of the two genes encoding mouse insulin, the insulin receptor in β cells, or the IGF-1 receptor in β cells in mice does not affect neonatal replication or accrual of β-cell mass [68]. Curiously, key cell cycle activators (Cyclins D1 and D2 and Cdk4) also appear dispensable for neonatal β-cell replication, but not for maintenance of β-cell mass in adulthood [69]. These results collectively suggest that the early signals driving replication of β cells during embryogenesis differs from signals that drive accrual of β-cell mass in early life (see later). The physiologic significance of the high neonatal β-cell turnover is a matter largely of speculation. Considering that neonatal islets show diminished or absent responsiveness to glucose-stimulated insulin secretion suggests that the neonatal turnover may be important for the eventual refinement and maturation of β cells [62]. Intriguingly, it has been suggested that this early turnover of β cells may result in exposure of β-cell autoantigens and trigger the pathogenesis of type 1 diabetes in susceptible individuals [70], although more recently this hypothesis has been challenged in mouse models [71].

**β-Cell growth with aging**

For many years it was speculated that β cells, much like neurons, were postmitotic and that their turnover in the mammal was minimal or zero. Over the last two decades, studies in mice have suggested a more dynamic picture, wherein β-cell mass can change in response to physiologic states such as growth, pregnancy, and obesity [72]. Following weaning in rodents, there is considerable increase in β-cell mass that reflects the increase in body mass and an adaptation to the needs for increased insulin release. β-Cell mass is reflective of the changes in β-cell formation, individual β-cell size, and β-cell death. Whereas replication can be fairly reliably estimated using thymidine analog incorporation strategies, the rate of death is much more difficult to measure, primarily because dying β cells are cleared from the islet rapidly and therefore are difficult to detect. Thus, in studies of mature rodents, the turnover of β cells is estimated in large part from rates of replication. Studies in rats have shown that β-cell mass increases in a near-linear fashion with body weight [63,73]. The replication rate of β cells declines with age (to ≤2% per day in 20 month-old animals), but does not approach zero, and β-cell volume increases with age. These results suggest a low, but clearly measurable, turnover of β cells in adult rats and a β-cell lifespan in the order of 1–3 months. Using a continuous BrdU labeling strategy in mice, a much lower rate of replication has been estimated in adult mice, leading to the conclusion that β-cell turnover is near zero [72]. Human β-cell mass accrual and replication rates are significantly more difficult to estimate, largely because cross-sectional data from a genetically diverse population must be extrapolated using static markers of replication (e.g. Ki67 or BrdU). Nevertheless, data from human autopsy samples suggest that there is accrual of β-cell mass with increasing body mass in children, with progressively decreasing potential for replication with age [66,74].

**Compensatory β-cell growth: adaptation to demand**

The concurrent growth of β-cell mass with growth of the organism is but one example of the capacity of β cells to adapt to the increasing metabolic demands of peripheral tissue. Physiologic states of tissue resistance to insulin action, such as obesity and pregnancy, pose similar challenges to the β cell. It was recognized as early as the 1930s from autopsy studies that the average size of the islets of Langerhans increases as humans become overweight [75]. Similar findings are seen in a variety of mouse, rat, pig, and other animal models of obesity, where both β-cell size and number are reportedly increased. The increase in β-cell mass in response to obesity reflects an adaptation to the increased insulin demands imposed by the resistance to insulin action in liver, muscle, and fat. Pregnancy
imposes a challenge on the β cell similar to obesity, as pregnancy causes a state of tissue resistance to insulin. In either obesity or pregnancy, inherent defects that prevent increases in β-cell mass and insulin release may be the underlying causes for the development of diabetes. For example, in mouse models haploinsufficient for the gene encoding Pdx1, there is impaired compensation for insulin resistance in terms of both β-cell mass and function, with ensuing glucose intolerance and diabetes [76]. Similarly, humans with heterozygous mutations of Pdx1 (a disorder known as maturity-onset diabetes of the young 4, or MODY4) develop diabetes with age, typically in adolescence or early adulthood [54]. In these individuals, it is thought that β-cell compensation for linear growth and/or age-related insulin resistance is impaired. Pdx1 is crucial not only in the regulation of genes encoding β-cell proteins that are important in insulin secretion, such as the glucose transporter Glut2, glucokinase, and insulin, but also in the regulation of genes that are downstream of the growth-promoting insulin receptor/insulin-like growth factor 1 (IGF-1) receptor signaling cascade [51].

**Origins of new β cells in the adult: neogenesis, transdifferentiation, and replication**

Considering the relatively small mass of β cells with respect to overall body mass, there has been a vigorous attempt over the last decade to define better the potential sources of new β cells in the growing mammal and to harness such sources for the creation of new β cells for those who are deficient. As discussed in the foregoing section, new β-cell formation was largely estimated by rates of β-cell replication, but excluded potential contribution from neogenesis. Therefore, if neogenesis were a major contributor to new β-cell formation, then rates of β-cell turnover were substantially underestimated. Speculation that a precursor β cell, or a true MPC, exists in the pancreas arose from early observations in rat models that new insulin-positive cells emanated from cells within proliferating [77]. The question of the origin of new β cells in models of pancreas regeneration has been addressed using lineage tracing analysis in mice to show that β cells arise almost exclusively by replication of preexisting insulin-positive cells rather than via neogenesis [78], a finding confirmed in subsequent studies in mice [79]. However, these findings do not exclude the possibility that a rare, insulin-positive cell type with high proliferative capacity (i.e., a cell type that would not be defined as a mature β cell) has the ability to serve as a MPC, or that under certain conditions other cell types within the pancreas (i.e., facultative stem cells) have the capacity to differentiate to β cells. Thus, investigators continue to posit the existence of these alternative cell types in the pancreas whose differentiation into mature β cells may recapitulate a pathway of transcription factor expression similar to that seen in development [80].

Because all pancreatic epithelial cell types arise from a common Pdx1-positive precursor, it has been proposed that mature pancreatic cells of either exocrine or endocrine origin may have the capacity to directly differentiate into β cells without the need for de-differentiation into a precursor form (a process known as “transdifferentiation”). In this respect, although lineage tracing analyses have all but ruled out the possibility that mature acinar cells transdifferentiate under normal conditions to β cells in mice [81], the ectopic expression of the key developmental transcription factors Pdx1, Neurog3, and MafA in acinar cells enables a program that allows their conversion to insulin-expressing cells [59]. Similarly, under specific experimental conditions in mice, mature α cells have the capacity to transdifferentiate into β cells [48]. Taken together, these studies reinforce the theme that different mature cell types of the pancreas that arise from a common origin have the capacity to exhibit phenotypic characteristics of one another, and leave open the possibility that under specific conditions such cell types may transdifferentiate to offset loss of β-cell mass.

Whether any of the mechanisms discussed earlier — neogenesis or transdifferentiation — play a role in human β-cell replenishment remains uncertain. Interestingly, studies in vitro suggest the potential existence of precursor cell types in the human pancreas [80], but it is unknown whether and to what extent such cells give rise to β cells normally in humans. To date, the best available data indicate that replication of preexisting β cells is the likely mechanism for accrual of β-cell mass during human growth [66,74,82].

**Regulators of β-cell growth: growth factors and cell cycle regulators**

A host of circulating factors appears crucial in the stimulation of early postnatal β-cell growth. As the growth of β cells closely parallels the growth of the organism during this early phase, it is relevant to note that nutrients, particularly glucose, remain among the most important factors contributing to β-cell replication during this period. Thus, intravenous glucose infusions for even short time periods (96 hours), which only mildly increase serum glucose concentrations, result in fivefold increases in β-cell replication in young mice [83]. Although an effect of glucose to directly stimulate β-cell replication has been proposed, it is possible that its effect may be caused by its stimulation of insulin release from β cells, such that insulin in an autocrine manner serves as the mitogen. Insulin and insulin-like growth factor 1 are classic growth factors that signal through related transmembrane receptors with associated receptor tyrosine kinases. Mice lacking the insulin receptor in β cells display impaired insulin release associated with reduced β-cell mass, whereas mice lacking the IGF-1 receptor in β cells display impaired insulin release without associated loss of β-cell mass. Interestingly, loss of both the insulin receptor and IGF-1 receptor in β cells results in severe reductions in β-cell mass and frank diabetes [68]. These data suggest that the insulin and IGF-1 signaling pathways function in distinct, but complementary ways, notwithstanding that both receptors share similar downstream signaling molecules (insulin receptor substrate proteins, phosphatidyl inositol-3 kinase, and protein...
kinase B). In recent years, a host of other growth factors has also been shown to positively influence β-cell replication and/or function (see Table 4.3), and include factors released not only from the islet, but also from a variety of organs, such as bone (osteocalcin), the anterior pituitary (growth hormone, prolactin), gut (glucagon-like peptide 1), fat (leptin, adiponectin), and brain (serotonin). Whereas these metabolites and growth factors can directly or indirectly impact β-cell replication and/or β-cell mass, it should be noted that their effects are much greater in younger mice and humans and much less so as aging occurs.

Although the effects of the aforementioned growth factors result in enhanced β-cell replication and insulin release, the pathways leading to activation of cellular replication machinery differ depending upon the factor [68]. Nevertheless, all factors ultimately impinge upon the components of the cell cycle. Transit through the cell cycle requires the β cell to exit the resting state (G0) and traverse G1, S, G2, and M states [82]. For the most part, replication of β cells is largely driven by factors that control the G1/S transition of the cell cycle. Genetic manipulation studies in mice have emphasized the importance of not only activators of the G1/S transition, but also inhibitors, such that the balance between the two appears to regulate the overall drive for β-cell replication. Cyclins and cyclin-dependent kinases (Cdks proteins) are major activators of β-cell replication. Cyclins and Cdks negatively regulate the major pocket protein known as pRb, which functions as a “molecular brake” on the G1/S transition. Cyclins and Cdks appear crucial in the accrual of early postnatal β-cell mass, but interestingly not in the generation of β-cell mass in the embryo. Mice homozygous null for the gene encoding CyclinD2 or Cdk4 exhibit no alterations in β-cell mass at birth, but show loss of mass accrual with age [84,85]. Similarly, loss of the gene encoding CyclinD1 does not affect embryogenesis, but heterozygosity of the CyclinD1 gene in combination with homozygous loss of CyclinD2 results in even further loss of β-cell mass with age and severe, life-threatening diabetes [84]. The cyclin-dependent kinase inhibitors (CKIs) — including p15Ink4b, p16Ink4a, p18Ink4c, p19Ink4d, p21Cip, p27Kip1, and p57Kip2 — are major negative regulators of β-cell proliferation, and their actions appear to predominate in later life, where these factors may be responsible for inhibition of β-cell proliferation in aging mammals [82].

Recent studies have clarified the human islet G1/S cell cycle protein expression pattern [86]. Whereas murine and human islets differ in their expression of the G1/S cell cycle activator Cdk4 (humans express Cdk6), they have virtually all G1/S CKIs in common. This latter observation may be crucial in the understanding of why β-cell replication is so dramatically reduced in aging humans. A particularly intriguing target in this respect is p16Ink4a, whose expression in β cells is up-regulated as mice age, and may serve as a target to prevent the age-induced limitations in β-cell mass [87]. The potential for β cells to undergo uncontrolled replication as a result of deregulation of G1/S cell cycle proteins is dramatically emphasized by mutations in the gene encoding Menin in both mice and humans. Menin is a tumor suppressor transcription factor that negatively regulates the expression of p18Ink4c and p27Kip1, and its absence or mutation results in the tumorigenic transformation of a variety of endocrine tissues (including β cells) in a syndrome known as multiple endocrine neoplasia 1 (MEN1) [88].

Conclusions and areas of future study

In the early twentieth century, the discovery of insulin dramatically transformed the treatment of diabetes mellitus. Indeed, it was thought that the administration of insulin might reduce stress and allow for the time necessary to regrow new β cells, a consequence that was never observed. In the ensuing decades, the incidence of type 2 diabetes rose to dramatic proportions, and as a result the quest for β-cell-based therapies for diabetes has seen broader appeal. As discussed, more recent research has led to dramatic insights into pancreas and β-cell development, and into the postnatal life cycle of the β cell. Although most of these insights derive from studies in lower animal species, their applicability to the treatment of human diabetes mellitus has risen to the forefront of discussion in recent years. Importantly, we know now that although β cells have the capacity to expand in the postnatal period, in humans the window for such expansion may be limited to the first 2–3 decades of life, and thereafter the ability to compensate for
physiologic stressors (such as obesity) diminishes with age. As such, strategies for therapies for diabetes in the future may well focus on ways to enhance β-cell replication or to engineer new β cells. With respect to the latter, studies of embryonic development have enabled important strides in generating β-like cells from primitive biologic precursors (e.g. human embryonic stem and induced pluripotent cells). Yet, these engineered cells do not exhibit the full phenotypic spectrum of true β cells, such as the ability to release insulin in response to a physiologic glucose challenge. The knowledge that all cells of the pancreas arise from a common progenitor has raised awareness that plasticity of fully differentiated pancreatic cell types may be much greater than originally thought. In this respect, studies of transdifferentiation of other abundant pancreatic cell types (such as α cells or acinar cells) to β cells in vivo may hold promise for the treatment of human diabetes, but to date no clear examples of human cell transdifferentiation have emerged. As the burden of diabetes increases, the need to translate research from lower animals to humans increases, and in the coming years it is likely that the generation of better model systems that mimic the human condition will become a greater priority.

References


CHAPTER 5

Pancreatic morphology in normal and diabetic states

Rebecca L. Hull and Steven E. Kahn
Division of Metabolism, Endocrinology and Nutrition, Department of Medicine, VA Puget Sound Health Care System and University of Washington, Seattle, WA, USA

Key points

- The pancreas comprises exocrine and endocrine compartments, which produce and release digestive enzymes and glucoregulatory hormones, respectively.
- The pancreatic islet is predominantly composed of endocrine (β, α, δ, F, and ε) cells, but is also supplied by an extensive capillary network and receives both parasympathetic and sympathetic innervation.
- The morphologic arrangement of these numerous cell types within the islet, and intercellular communication between them appears to be critical for normal endocrine function.
- The islet extracellular matrix is an emerging site of regulation of β-cell function and survival, and is disrupted or degraded in diabetes.
- In type 1 diabetes, an almost complete obliteration of β cells occurs, predominantly via T-cell-mediated autoimmune destruction.
- In type 2 diabetes, pancreas weight is relatively normal, but loss of β cells also occurs, albeit to a lesser extent than in type 1 diabetes.
- β-Cell loss also occurs in cystic fibrosis-related diabetes, suggesting that exocrine pancreatic abnormalities can have profound effects on the islet.
- In type 2 diabetes, the etiology of β-cell demise is complex. Nutrient excess, amyloid deposition and inflammation have all been proposed as underlying mechanisms of β-cell loss.

Pancreatic anatomy and morphology

The pancreas is located in the upper abdominal cavity, in close proximity to both the duodenum and spleen (Figure 5.1(a)). In humans, pancreatic weight ranges from 40 – 150 g [1,2]. Embryonically, the pancreas develops from two separate buds of the primitive foregut, yielding duodenal (ventral) and splenic (dorsal) lobes [3]. Once developed, these constitute the head and body/tail, respectively. These regions of the pancreas are similar, aside from some differences in the distribution and composition of pancreatic islets (discussed later).

Functionally, the pancreas comprises two independent compartments, the exocrine and endocrine pancreas, which derive from common endodermal precursor cells during development [3]. The exocrine pancreas accounts for the vast majority of pancreatic mass and consists of lobules, each comprising acini that connect into a network of ducts (Figure 5.1(b)). Acinar cells demonstrate a characteristic morphology at the electron microscopy level including electron-dense zymogen granules and an extensive endoplasmic reticulum (ER) network. This ultrastructural organization is consistent with the chief function of the exocrine pancreas, namely to secrete digestive enzymes including amylase and trypsin via the pancreatic ductal system into the gut. The endocrine pancreas comprises islets of Langerhans, roughly spherical structures that contain hormone-producing cells. Islets constitute only a small minority of pancreatic mass but are critically important for metabolism throughout the body, and especially for maintenance of glucose homeostasis. This was first shown in 1890 by studies in which pancreatectomy in dogs resulted in diabetes [4]. In the human pancreas, it has been estimated that there are one to two million islets scattered throughout the exocrine pancreas, which together comprise approximately 2% of pancreatic mass [5,6]. While the function of these two pancreatic compartments differ significantly, they exist in close proximity to one another, and evidence exists for interactions between them. Specifically, islets are distributed throughout the exocrine pancreas. Islet hormone-rich blood perfuses the exocrine pancreas [7], and the islet hormones, insulin and pancreatic polypeptide, have been shown to stimulate amylase secretion from the exocrine pancreas, while glucagon inhibits amylase secretion [8–10]. This suggests that the functional status and viability of the islet can influence function/viability of the exocrine pancreas and potentially vice versa.
Islet composition and morphology

Despite the relatively small size of pancreatic islets, each comprising approximately 2000–4000 cells [11], every one is a complex mini-organ, containing numerous cell types (Figure 5.1(c)). Endocrine cells are, not surprisingly, the most abundant islet cell type and are characterized at the ultrastructural level by the presence of numerous secretory granules that contain endocrine hormones packaged and stored ready for immediate release in response to the appropriate stimulus. Additionally, endocrine cells contain a high density of mitochondria and abundant ER (although not so dense as the surrounding acinar cells). These endocrine cells are subcategorized based on their predominant hormone constituent, as follows.

Insulin-producing $\beta$ cells are the predominant islet cell type. They were first identified in 1907 by silver staining [12] and were the second islet endocrine cell type to be described. $\beta$ Cells comprise 50–80% of islet volume, depending on the species in question; rodent islets typically have a higher proportion of $\beta$ cells (around 60–80%) [13,14], while nonhuman primate and human islets have a lower relative proportion (around 50–60%) [13–19] (Figure 5.2). The distribution of $\beta$ cells within islets also differs among species. In rodents $\beta$ cells are predominantly located in the core of the islet [13,14,19], while in nonhuman primate and human islets they are more evenly distributed throughout the islet (Figure 5.1(c)) [13,14,18]. Regardless of their location within the islet, for all species $\beta$-cell secretory granules have a characteristic appearance at the electron microscope level, with an electron-dense core that arises due to the formation of insulin hexamers which are cross-linked with zinc [11]. This granule core is surrounded by an electron-lucent “halo” [11]. Release of insulin from $\beta$ cells is critically important in regulating blood glucose levels, predominantly acting to suppress hepatic glucose production and to enhance glucose uptake in insulin-sensitive tissues such as skeletal muscle and adipose tissue. Mechanisms of insulin release are discussed in detail in Chapters 7, 8, and 9. In addition to insulin, $\beta$ cells also produce islet amyloid polypeptide (IAPP), which is co-localized with insulin in secretory granules and is therefore co-secreted with insulin [20].

Glucagon-producing $\alpha$ cells were actually the first islet endocrine cell type to be described and are so called due to the alcohol fixation (“A”-cells) that was used to identify them [12].
These form the second most abundant endocrine cell type in the islet, but are far less abundant than β cells. In rodent islets, they account for around 15% of islet area [19], whereas in humans they are more abundant (around 35%; [13]) (Figure 5.2). Also, like β cells, α cells differ in their distribution within islets across species. In rodents, they are located almost exclusively in the mantle of the islet [13,14,19], while in human and nonhuman primate islets they are more evenly distributed throughout the islet (Figure 5.1(c)) [13,14,18]. Unlike β cells, whose abundance in islets is relatively constant throughout the pancreas, islets in the head of the pancreas are richer in α cells than those in the tail [19]. Glucagon-containing secretory granules differ in their appearance compared to insulin-containing granules, in that they lack an electron lucent halo. Thus, the entire secretory granule in the α cell is electron dense [21]. The hormonal effects of glucagon act to oppose those of insulin, stimulating hepatic glucose production, and being a critical mediator of the restoration of normal glucose levels following hypoglycemia. Glucagon secretion and its effects are described in Chapter 10.

Somatostatin-producing δ cells comprise the third islet endocrine cell type, being lower in abundance (<10% of islet cells) than either β or α cells (Figure 5.2). Their distribution within islets is similar to that of α cells, while in primates and humans they are more evenly distributed throughout [13,14,18]. While few studies have systematically assessed δ-cell distribution in islets throughout the pancreas, δ-cell number does not appear to be markedly altered in islets from different regions of the pancreas [19]. Ultrastructurally, somatostatin-containing granules are homogeneous in appearance, similar to α-cell granules, but slightly less dense. The effects of islet-derived somatostatin are thought to be largely paracrine in nature due to its short half-life [22]. Somatostatin acts on receptors present on β and α cells, inhibiting both insulin and glucagon release.

Pancreatic polypeptide-producing F cells are less abundant still than δ cells, contributing on average <5% of islet cells. Again like α cells, their distribution differs throughout the pancreas, with islets in the head of the pancreas containing most of the pancreatic polypeptide-producing F cells [19,23]. Thus, islets in this region of the pancreas can contain up to 20% pancreatic polypeptide-containing cells, at the expense of α cells, which are decreased in this region. Conversely, in the rest of the pancreas, F cells only account for 1–3% of islet cells [19,23] (Figure 5.2). F-cell granules are small, but similar in appearance to β-cell granules, containing an electron-dense core and electron-lucent halo. The release of pancreatic polypeptide from islets serves as an indicator of vagal outflow [24]. However, the function of pancreatic polypeptide, once released, is unclear. Some studies have suggested a role in inhibition of pancreatic enzyme and bile secretion, while others have suggested a role in modulation of food intake and energy expenditure [9,25].
first [33]. Thus, α cells are exposed to high concentrations of insulin as they lie downstream of β cells in the islet vasculature, which tonically inhibit glucagon secretion. δ Cells are thought to be downstream of both β and α cells, consistent with the notion that somatostatin’s effects to inhibit insulin and glucagon secretion occur in a paracrine, rather than endocrine manner. While the distribution of endocrine cells within nonhuman primate or human islets differs from that of rodents, it is likely that the same directional vascular supply exists [34,35].

Interestingly, there are differences in islet vascular density among species. Rodent islets contain a dense network of small capillaries [30,31], while human islets appear to contain fewer, larger capillaries (Hull, Brissova, Powers, unpublished observation) [36]. However, the functional consequence of this difference in capillary density is unknown. Islet capillaries are lined by a highly fenestrated endothelium, with islet endothelial cells containing around 10 times more fenestrae than capillaries in the neighboring exocrine pancreas [37]. This fenestration allows rapid exchange of nutrients and oxygen between blood and islet cells. However, in contrast to the liver which contains open fenestrae in its endothelium, islet endothelial fenestrae are gated; that is, covered by a glyocalyx, a semipermeable layer composed predominantly of the polysaccharide heparan sulfate [38]. This suggests some selectivity exists with respect to the molecules that can readily pass in and out of the islet capillary, although this is poorly understood. While the islet vasculature is critically important for providing adequate blood flow, supplying nutrients to islet cells and facilitating delivery of islet hormones to peripheral tissues, islet capillaries also provide important signals for normal islet endocrine growth and survival [39–41]. Finally, a vascular basement membrane, a specialized form of extracellular matrix, exists between islet capillaries and endocrine cells [36,37,40,42–44]. While this extracellular matrix comprises predominantly collagen IV and laminins, it also contains a complex array of other proteins and proteoglycans including heparan sulfate proteoglycans, nidogens and hyaluronan which provide both structural support along with critical signals to both islet endothelial cells and endocrine cells, and participates in maintenance of normal function and proliferation of islet cells [40,45,46].

**Islet innervation.** The islet receives extensive autonomic input, via both sympathetic and parasympathetic branches of the autonomic nervous system [47]. These nerves do not form classical synapses with islet endocrine (or other) cells, but form terminals that release neurotransmitters in close proximity to islet cells which in turn act as important regulators of islet endocrine hormone release [47]. Islet innervation and its functional consequences are reviewed in detail in Chapter 9. Morphologically, islet innervation mirrors that of vascularization, with nerve fibers running parallel to islet capillaries [48]. Accordingly, rodent islets containing numerous, fine nerve fibers [48], while in contrast, human islets contain fewer, larger nerve fibers [49] (Figure 5.4).

**Interactions among islet cell types.** This highly ordered distribution of islet cell types has functional consequences. Islet cell types interact with one another by a number of different mechanisms including direct cell–cell contact, release of paracrine signals or via the extracellular matrix. For example, signaling via gap junctions is important for coordinating insulin release [50], while autocrine and paracrine signals such as GABA or somatostatin can enhance or suppress islet hormone release from neighboring endocrine cells [22,51,52]. Exposure of islets to neurotransmitters (reviewed in Chapter 9) or endothelial-derived factors such as hepatocyte growth factor, laminin or thrombospondin-1 can stimulate β-cell secretory

---

**Figure 5.4** Innervation of a mouse (a) and human (b) islet, visualized using the axonal marker synapsin I/II. Source: Adapted from Rodriguez-Diaz 2011 [49]. Reproduced with permission of Elsevier.
function and/or replication [41,45,53]. Conversely, β cells produce factors such as vascular endothelial growth factor, which are essential for islet endothelial cell viability and function [40,41]. Culture of β cells on extracellular matrix has profound effects to enhance islet cell proliferation, survival and function, suggesting another mechanism by which endocrine cells can be influenced by the islet vasculature [40,54]. Thus, changes in the abundance or organization of any one of the multiple islet cell types, or in exocrine pancreatic viability and function, likely has significant consequences for islet health and function and ultimately for glucose homeostasis.

Islet β cell regenerative potential

As discussed in Chapter 4, development and organization of islet endocrine cells occurs during embryogenesis and the early post natal period [3]. In rodents, continued β-cell expansion can occur into adulthood [55]. However, the ability of rodent β cells to replicate is severely limited with increasing age [56]. In humans, there is some plasticity in islet volume, particularly in very young individuals [57], but there is very limited regenerative potential of β cells in adults [57,58]. Physiologic stimuli such as insulin resistance, obesity, and pregnancy have been reported to result in increased β-cell mass in humans and rodents [59–62]; however, pancreatectomy, which is known to stimulate β-cell regeneration in young rodents [56,63], does not result in increased β-cell replication in older rodents [56] or adult humans [64]. Altering islet volume appears to be an exquisitely regulated process; in line with the requirement for maintenance of the organization of islet cell types, β-cell replication in rodents during pregnancy is preceded by islet angiogenesis (expansion of islet capillaries), suggesting that expansion of the islet vascular supply is required to allow expansion of the endocrine cell population [41]. Maintenance of the appropriate proportions and organization of islet endocrine cell populations in the face of islet expansion also occurs in response to high fat feeding in mice [65]. Islet innervation is also increased under conditions of high fat feeding and insulin resistance [66,67]. However, despite the ability of the islet cell population to expand in response to some physiologic stimuli, the limited capacity for β-cell expansion, especially with age, becomes a major problem in disease states where β cells are targeted for destruction.

Disturbances in pancreas/islet morphology in diabetes

Type 1 diabetes

Type 1 diabetes (T1DM) is classically associated with autoimmune destruction of β cells [68] (Figure 5.2). However, the pancreas is more broadly affected, with overall pancreas size being decreased in individuals with this form of diabetes [68,69], and loss of exocrine tissue occurring close to areas of immune infiltration [70]. β-Cell destruction is largely a T-cell-mediated process, involving mainly CD8+ cells, but also including CD4+ cells and other immune cells such as macrophages and B cells [71,72]. Lymphocytic infiltration of islets is well documented in animal models [73,74]. However, the degree of infiltration can vary widely even among islets from the same animal in various stages of diabetes development [73,74]. Further, the extent of leukocyte infiltration in humans appears to be less than that seen in animal models, while the variability in affected islets is similar [70,72,75]. In human T1DM, insulitis is primarily reported in individuals with recent onset disease [70,76], although it has been detected in patients 8 years following diagnosis [72,76]. That insulitis occurs predominantly around the time of disease onset is consistent with the clinical observation that the largest decline in C-peptide responses occurs between 6 months prior to and 12 months following disease diagnosis [77,78]. Despite the variability in detectable insulitis, autoimmune destruction appears to result in eventual elimination of the majority of β cells [70,79]. However, β cells can persist for many years into the course of the disease [79,80] and low levels of β-cell replication have been documented in some [81], but not all studies [82]. Further, there is evidence for residual insulin release many years after the development of hyperglycemia [83]. This raises the possibility that β-cell destruction may not be complete and that regeneration may be possible. A recent study documenting the efficacy of stem cell therapy in rapidly reversing T1DM, in many cases up to 36 months of follow-up, provides support for this concept [84].

While β-cell destruction is widespread in T1DM, non-β-cell islet populations, particularly α cells, appear to be spared the autoimmune destruction [70]. However, despite the persistence of α cells in T1DM, their function is undoubtedly dysregulated. Specifically, meal-stimulated glucagon responses are exaggerated [85], while glucagon release in response to hypoglycemia is markedly impaired [86]. These abnormalities may be due to the lack of oscillating insulin levels, which would normally act to regulate glucagon release [87]. However, the lack of glucagon response to hypoglycemia is likely also impacted by the early loss of sympathetic nerve terminals in islets, which has been demonstrated in rodent models of T1DM [88,89] and in human T1DM [90]. This islet neuropathy is selective, with islet parasympathetic innervation appearing to be normal, at least in rodent models of T1DM [90].

Whether islet capillary density is altered in T1DM is currently unknown. Recently however, significant alterations in the extracellular matrix closely apposed to the islet vasculature have been described in human T1DM and animal models thereof [36,42]. Degradation of peri-islet extracellular matrix has been shown to correlate with leukocyte infiltration and β-cell loss in human T1DM and NOD diabetic mice [36,42]. Interestingly, however, once insulitis is resolved, peri-islet extracellular matrix is regenerated, even in the absence of insulin-positive cells, providing further support for a role of leukocytic infiltration in
the degradation of this extracellular matrix. Altered localization of the extracellular matrix component hyaluronan [91,92] and increased production of extracellular matrix degrading enzyme heparanase [93] have also been described in association with lymphocytic infiltration of islets in NOD diabetic mice, in common with other autoimmune diseases [94]. The contribution of these changes in extracellular matrix to diabetes onset and progression are not fully understood at present, but they may be important in allowing leukocytes to gain access to the islet, and are an active area of investigation.

**Type 2 diabetes**

Macroscopically, the pancreas appears largely unchanged in T2DM. Fibrosis in the exocrine pancreas has been described [95], suggesting some abnormality in the exocrine pancreas, but this has not been widely studied. In contrast, the presence of morphologic abnormalities in islets from subjects with T2DM has long been established. More than a century ago, Opie described decreased cell number and accumulation of what was later identified as islet amyloid [96]. Subsequently, it was confirmed that islet \( \beta \)-cell volume is decreased in T2DM [59,97], an observation has been reproduced in numerous studies, across several ethnic groups [15–17] (Figure 5.2). Butler et al. additionally showed that \( \beta \)-cell volume is also decreased in subjects with impaired fasting glucose, with the extent of reduction being intermediate between that of subjects with T2DM and nondiabetic controls [15]. Overall, the extent of \( \beta \)-cell loss reported varies widely among studies (0–63% reduction), most likely due to the variability of \( \beta \)-cell volume among subjects [16,17] and also to the site of sampling [16]. Similar to the situation in T1DM, islet \( \alpha \)-cell mass has been shown to be maintained in T2DM, resulting in a relative increase in the \( \alpha:\beta \) cell ratio [2,95,98]. In animal models, islet glucagon and pancreatic polypeptide immunoreactivity have been reported to be similar or increased relative to nondiabetic animals [99,100], while somatostatin immunoreactivity is more variable, being reportedly increased, similar or decreased in comparison to nondiabetic animals [99–101].

Alterations in density and/or morphology of islet capillaries have been described in a variety of rodent models of diabetes. Early in the course of hyperglycemia, distorted islet capillary morphology is present and with more advanced diabetes, loss of capillary density occurs and is frequently associated with islet fibrosis [102–108]. No published studies have been performed on human pancreas specimens, but our unpublished data suggest that while islet capillary morphology is distorted, islet capillary density is not decreased in T2DM relative to nondiabetic controls (Brissova, Powers, Hull, unpublished observation). Decreased islet innervation has also been reported in animal models of T2DM [109], but has not been determined in humans with the disease. Abnormalities in islet extracellular matrix have also been documented in human T2DM and animal models thereof. These include accumulation of islet amyloid, which comprises the aggregated form of the \( \beta \)-cell peptide IAPP [17,95,110], and islet fibrosis occurring due to fibrillar collagen deposition [111,112].

**Influence of exocrine pancreas abnormalities on islet morphology and function**

Diseases affecting the exocrine pancreas are associated with diabetes. Acute pancreatitis has been associated with glucose intolerance and impaired insulin release, but this disturbance seems to be temporary [113], suggesting that exocrine pancreas abnormalities can impact islet function. In cases of chronic pancreatitis whose primary disease etiology is exocrine in nature, diabetes is present in the majority of cases [114]. However, pancreatitis is also more common in individuals with T2DM [95,115], making the link between exocrine disease and the subsequent onset of diabetes less clear.

Cystic fibrosis is an autosomal recessive disorder, arising due to one of several mutations in the cystic fibrosis transmembrane receptor (CFTR), a chloride channel, with disease onset usually occurring in childhood [116]. Lung disease is the primary manifestation of CFTR mutation. However, with improved treatment including lung transplantation, survival has significantly improved in recent years; as a result, other complications of cystic fibrosis are now more common. Pancreatic involvement, namely significant exocrine pancreas fibrosis is the second most common feature of cystic fibrosis, after lung pathology. Accordingly, cystic fibrosis-related diabetes complicates a large proportion of cystic fibrosis cases [117]. This form of diabetes does not seem to include underlying autoimmunity, suggesting its etiology differs from that of T1DM [118]. Unlike T2DM, insulin resistance does not appear to be a major underlying cause [119,120]. However, defective insulin release has been clearly demonstrated [119,120]. This is accompanied by decreased islet \( \beta \)-cell volume, which has been documented in several studies [121,122]. The mechanisms of \( \beta \)-cell loss in cystic fibrosis-related diabetes remain unclear, although islet amyloid deposition is also present in this population [123], suggesting that at least certain aspects of islet pathology share features with T2DM. Thus, some mechanisms that may explain \( \beta \)-cell loss in T2DM, and which are discussed later, are likely also pertinent to cystic fibrosis-related diabetes.

**Mechanisms of \( \beta \)-cell loss in type 2 diabetes**

While alterations in several islet cell types have been reported in T2DM or animal models thereof, only \( \beta \) cells have reproducibly been shown to be reduced [15–17,59,96,97]. Decreased \( \beta \)-cell volume in T2DM is associated with an increase in \( \beta \)-cell apoptosis [15,17], which occurs without a compensatory increase in \( \beta \)-cell replication due at least in part to the limited regenerative capacity of adult human islets [57,58]. Thus, mechanisms that result in \( \beta \)-cell apoptosis or other forms of \( \beta \)-cell death appear
to be critical for loss of β cells in T2DM. This process has been widely studied, and numerous mechanisms have been implicated.

**Chronically elevated glucose and/or free fatty acids**

Type 2 diabetes is characterized by increased circulating nutrients including glucose and free fatty acids (FFA). The literature clearly shows that chronic exposure of β cells to elevated glucose results in impaired β-cell function [124–126], but the data regarding cellular toxicity in response to this nutrient are more mixed. Exposure of cultured β-cell lines or islets to high glucose can, in some cases, result in increased β-cell death [127–132]. This may result from oxidative stress, activation of Fas receptor-mediated or mitochondrial apoptosis and may involve thioredoxin interacting protein (TXNIP) [127–129]. However, this effect of glucose to induce apoptosis is not a universal finding; several studies have demonstrated that the “toxic” effects of chronic hyperglycemia to impair β-cell function are reversible even after several weeks in culture [133,134]. Further in vitro and in vivo studies exposing β cell to elevated glucose have shown beneficial effects with glucose-promoting survival signals, suppressing apoptosis [135] or resulting in increased β-cell replication [136–138].

Exposure of islets/β cells to increased FFA levels alone, or in the presence of hyperglycemia results in impaired insulin release [139–141]. Culture of β cells in the presence of increased FFA, particularly palmitate, can also result in β-cell apoptosis [132,142–144]. This has been shown to occur via some of the same mechanisms as glucose-induced apoptosis, namely oxidative stress, ER stress and activation of the mitochondrial apoptosis pathway, and additionally may require increased ceramide or nitric oxide levels. However, similar to the observations with elevated glucose, high fat feeding or lipid infusions in vivo result in increased β-cell mass, as a result of increased β-cell replication [62,138].

Thus, taken together, the effects of nutrient excess appear to be more detrimental to β-cell secretory function rather than clearly inducing β-cell death, and some of these effects may be reversible.

**Islet amyloid**

As mentioned, amyloid deposition occurs in islets in the majority of subjects with T2DM, as well as in subjects with cystic fibrosis-related diabetes [17,95,110,123]. Accumulation of islet amyloid occurs due to the aggregation of the normally soluble β-cell peptide IAPP, which is then deposited in the islet extracellular matrix, between islet capillaries and β cells. This aggregation only appears to occur under conditions of diabetes or islet dysfunction, with islet amyloid being relatively rare in individuals without diabetes even in individuals with extremely high circulating levels of IAPP [110,145,146]. The underlying cause of this IAPP aggregation is unclear, but may involve impaired processing of IAPP from its precursor proIAPP [147–149] and/or interaction between IAPP and extracellular matrix components, principally heparan sulfate proteoglycans [150–153]. Human autopsy studies have yielded somewhat conflicting results, but the literature clearly demonstrates that the extent of islet amyloid deposition is associated with decreased β-cell volume [17,154] and increased β-cell apoptosis [17] (Figure 5.5). Studies using cultured human islets and transgenic animals expressing human IAPP (mouse and rat IAPP are not amyloidogenic) have further elucidated the mechanism(s) by which IAPP aggregation may elicit β-cell toxicity. Culture of human or transgenic mouse islets under conditions that favor amyloid formation, for example high glucose, result in amyloid-induced oxidative stress and increased β-cell apoptosis, thereby leading to a reduction in β-cell area [155–160]. This β-cell loss can occur via activation of the cell surface death receptor Fas [161], or cJun N-terminal kinase (JNK) and downstream activation of apoptosis [162]. Additionally, when human IAPP aggregation and thereby amyloid formation is

![Figure 5.5](image_url)  
*Figure 5.5* Relationship between islet amyloid deposition and decreased β-cell area (a) \((r = -0.76, p < 0.001)\), and increased β-cell apoptosis (b) \((r = 0.56\) and \(p < 0.01)\) in human autopsy pancreas specimens from subjects with type 2 diabetes (circles) and nondiabetic controls (triangles). Source: Adapted from Jurgens 2011 [17]. Reproduced with permission of Elsevier.
inhibited by Congo red [157] or overexpression of the enzyme neprilysin [163] β-cell apoptosis is reduced, suggesting IAPP aggregation is an important mediator of β-cell toxicity. Some, but not all, studies have demonstrated that expression of human IAPP results in an ER stress response [164–166]. However, this appears to be related to the magnitude of IAPP overexpression, and does not occur at physiologic levels of human IAPP, nor does it differ between human islets from individuals with T2DM who do or do not have amyloid deposits [166]. Finally, recent data have shown that human IAPP in its aggregated form is proinflammatory, eliciting cytokine and chemokine production from macrophages/dendritic cells [167,168]. Further, islet IL-1β expression may be increased in conditions of amyloid deposition [112,161,167], suggesting a novel mechanism by which islet amyloid may result in β-cell death.

**Islet inflammation**

As discussed earlier, inflammation in the islet has long been established as a hallmark of T1DM. Islet infiltration and release of molecules such as proinflammatory cytokines have clearly been implicated in β-cell death in this form of diabetes [169]. In T2DM, the concept that low-grade, chronic inflammation exists, most likely associated with insulin resistance, is a relatively new idea. As this field of research has emerged, so too has the hypothesis that inflammation in the islet may play a role in β-cell death in T2DM [112]. However, this remains a controversial area. Evidence in favor of a role for islet inflammation includes reports of increased islet production of interleukin 1β following chronic high glucose culture of human islets [127]. Islet interleukin 1β production has also been suggested in models of islet amyloid formation [112,161,167]. Activation of signaling pathways associated with the innate immune response (namely toll-like receptors) has been shown to occur in β cells in response to agents such as FFA or lipopolysaccharide [170,171]. This activation can lead to β-cell toxicity and death, suggesting that inflammation may play a role in the demise of the β cell in T2DM.

**Summary and future directions**

The morphology of the pancreas and pancreatic islet is complex, and disturbances in pancreas and islet volume/arrangement that occur in diabetes are multifactorial. Loss of β cells is a common feature of type 1-, type 2-, and cystic fibrosis-related diabetes. However, the mechanisms that underlie this pathology differ significantly among the various types of diabetes. Our understanding of how β-cell destruction occurs in type 1 and type 2 diabetes has been improved by a large number of studies, but we still have much to learn about how this occurs. Emerging areas of interest include understanding how changes in islet vasculature, innervation, and extracellular matrix contribute to derangements in islet morphology, which may in turn shed new light on the causes of β-cell loss in diabetes.

**Acknowledgments**

This work was supported by the Department of Veterans Affairs (Seattle Division VA Puget Sound Health Care System, Seattle, WA, USA) and National Institutes of Health grants DK088082 (RLH), DK017047 and DK075998 (SEK).

**References**


77 Greenbaum CJ, Beam CA, Boulware D, et al.: Fall in C-peptide during first 2 years from diagnosis: evidence of at least two distinct phases from composite Type 1 Diabetes TrialNet data. *Diabetes* 2012;61:2066–2073.
Diabetes 2011; Jour-
Diabetologia 1901; Diabetes
Diabetologia 1993; 2008;
2012; PNAS
106 Li X, Zhang L, Meshinchi S, et al.: Islet microvasculature in islet
105 Tikellis C, Wookey PJ, Candido R, et al.: Improved islet morphol-
104 Mizuno A, Noma Y, Kuwajima M, et al.: Changes in islet capil-
103 Nakamura M, Kitamura H, Konishi S, et al.: The endocrine pan-
100 Makino H, Matsushima Y, Kanatsuka A, et al.: Changes in pancre-
99 Howarth FC, Al Kitbi MK, Hameed RS, Adeghate E: Pancreatic
98 Yoon KH, Ko SH, Cho JH, et al.: Selective beta-cell loss and
97 Maclean N, Ogilvie RF: Quantitative estimation of the pancreatic
95 Clark A, Wells CA, Buley ID, et al.: Islet amyloid, increased A-cells,
94 Sorokin L: The impact of the extracellular matrix on inflammation.
89 Taborsky GJ, Jr, Mei Q, Hackney DJ, et al.: Loss of islet sympathetic
nervous system in mediating the glucagon response to hypo-
in non-obese diabetic mice by targeting CD44 with a specific mon-
86 Bollyky PL, Bogdani M, Bollyky JB, et al.: The role of hyaluronan
and the extracellular matrix in islet inflammation and immune reg-
83 Ziolkowski AF, Popp SK, Freeman C, et al.: Heparan sulfate and
heparanase play key roles in mouse beta cell survival and autoim-
82 Sorokin L: The impact of the extracellular matrix on inflammation. 
81 Clark A, Wells CA, Buley ID, et al.: Islet amyloid, increased A-cells,
reduced B-cells and exocrine fibrosis -- quantitative changes in the
80 Opie EL: The relation of diabetes mellitus to lesions of the pan-
icreas: hyaline degeneration of the islets of Langerhans. Journal of
79 Maclean N, Ogilvie RF: Quantitative estimation of the pancreatic
78 Yoon KH, Ko SH, Cho JH, et al.: Selective beta-cell loss and
alpha-cell expansion in patients with type 2 diabetes mellitu-
s in Korea. Journal of Clinical Endocrinology & Metabolism
2003;88:2300 – 2308.
77 Howarth FC, Al Kitbi MK, Hameed RS, Adeghate E: Pancreatic
peptides in young and elderly Zucker type 2 diabetic fatty rats. Jour-
76 Makino H, Matsushima Y, Kanatsuka A, et al.: Changes in pancre-
atic somatostatin content in spontaneously diabetic mice, as deter-
dined by radioimmunoassay and immunohistochemical methods.
75 Patel YC, Cameron DP, Stefan Y, et al.: Somatostatin: widespread
abnormality in tissues of spontaneously diabetic mice. Science
74 Like AA, Chick WL: Studies in the diabetic mutant mouse. II. Elec-
73 Nakamura M, Kitamura H, Konishi S, et al.: The endocrine pancre-
s of spontaneously diabetic db/db mice: microangiopathy as revealed
by transmission electron microscopy. Diabetes Research and Clinical
72 Mizuno A, Noma Y, Kuwajima M, et al.: Changes in islet capil-
larly angiarchitectue coincide with impaired B-cell function but
not with insulin resistance in male Otsuka-Long-Evans-Tokushima
fatty rats: dimorphism of the diabetic phenotype at an advanced
71 Tikellis C, Wokeye PJ, Candido R, et al.: Improved islet morphol-
ogy after blockade of the renin- angiotensin system in the ZDF rat.
70 Li X, Zhang L, Meshinchi S, et al.: Islet microvasculature in islet
hyperplasia and failure in a model of type 2 diabetes. Diabetes
69 Taborsky GJ, Jr, Mundinger TO: Minireview: The role of the auto-
nervous system in mediating the glucagon response to hypo-
in non-obese diabetic mice by targeting CD44 with a specific mon-
67 Bollyky PL, Bogdani M, Bollyky JB, et al.: The role of hyaluronan
and the extracellular matrix in islet inflammation and immune reg-
66 Clark A, Wells CA, Buley ID, et al.: Islet amyloid, increased A-cells,
reduced B-cells and exocrine fibrosis -- quantitative changes in the
65 Opie EL: The relation of diabetes mellitus to lesions of the pan-
icreas: hyaline degeneration of the islets of Langerhans. Journal of
64 Maclean N, Ogilvie RF: Quantitative estimation of the pancreatic
63 Yoon KH, Ko SH, Cho JH, et al.: Selective beta-cell loss and
alpha-cell expansion in patients with type 2 diabetes mellitu-
s in Korea. Journal of Clinical Endocrinology & Metabolism
2003;88:2300 – 2308.
62 Howarth FC, Al Kitbi MK, Hameed RS, Adeghate E: Pancreatic
peptides in young and elderly Zucker type 2 diabetic fatty rats. Jour-
61 Makino H, Matsushima Y, Kanatsuka A, et al.: Changes in pancre-
atic somatostatin content in spontaneously diabetic mice, as deter-
dined by radioimmunoassay and immunohistochemical methods.
60 Patel YC, Cameron DP, Stefan Y, et al.: Somatostatin: widespread
abnormality in tissues of spontaneously diabetic mice. Science
59 Like AA, Chick WL: Studies in the diabetic mutant mouse. II. Elec-
58 Nakamura M, Kitamura H, Konishi S, et al.: The endocrine pancre-
s of spontaneously diabetic db/db mice: microangiopathy as revealed
by transmission electron microscopy. Diabetes Research and Clinical
57 Mizuno A, Noma Y, Kuwajima M, et al.: Changes in islet capil-
larly angiarchitectue coincide with impaired B-cell function but
not with insulin resistance in male Otsuka-Long-Evans-Tokushima
fatty rats: dimorphism of the diabetic phenotype at an advanced
56 Tikellis C, Wokeye PJ, Candido R, et al.: Improved islet morphol-
ogy after blockade of the renin- angiotensin system in the ZDF rat.
55 Li X, Zhang L, Meshinchi S, et al.: Islet microvasculature in islet
hyperplasia and failure in a model of type 2 diabetes. Diabetes


156 Marzban L, Tomas A, Becker TC, et al.: Small interfering RNA-mediated suppression of proislet amyloid polypeptide expression


**Key points**

- Glucose stimulates *insulin* gene transcription and insulin mRNA stability.
- Glucose regulates the binding of key transcription factors to the *insulin* gene, notably Pdx-1, MafA, and NeuroD1/Beta2.
- Chronic hyperglycemia and dyslipidemia impair *insulin* gene expression via Pdx-1 and MafA.
- Glucose stimulates translation of the mRNA into proinsulin and processing of proinsulin into mature insulin.
- Proinsulin is processed in the immature granule compartment by successive cleavage at two dibasic sites by the prohormone convertases PC2 and PC1/3.
- Chronic nutrient excess and hyperinsulinemia in type 2 diabetes impose a high secretory demand to the β cell which is not adequately matched by an increase in biosynthetic capacity.

**Introduction**

The unique property of the pancreatic β cell is its ability to secrete insulin to enable circulating glucose levels to be maintained within a narrow physiologic range, despite wide fluctuations in energy intake and expenditure. It is able to sense the glucose concentration in the extracellular milieu, and adapt its insulin secretion rate via a complex interplay between nutrients, hormones, and neuronal signals. Whereas the minute-to-minute regulation of insulin secretion occurs at the level of exocytosis of pre-formed insulin, adaptation to long-term changes in the environment also involves regulated changes in the transcription rate of the *insulin* gene, translation of the mRNA, and processing of the proinsulin molecule into fully mature insulin. These processes are coordinately regulated by glucose (Figure 6.1) under normal circumstances, and their perturbation leads to β-cell dysfunction and type 2 diabetes.
Insulin gene expression and biosynthesis

Insulin gene expression is principally controlled by a highly conserved region lying approximately 340 bp upstream of the transcription initiation start, termed the enhancer/promoter control region [14,15]. Considerable progress has been made in defining the many different cis- and trans-acting factors that ensure precise transcriptional regulation, with the focus here on describing the β-cell-enriched transcription factors most pertinent to metabolically regulated expression, specifically Pdx-1, NeuroD1/Beta2, and MafA.

Pdx-1 is a homeodomain protein that plays a major role in pancreatic β-cell development and function [16,17]. It primarily binds as a monomer to the conserved AT-rich A3 box (-201/-196 bp) and activates insulin transcription, although this protein also appears to act as a repressor in other gene contexts [18]. Pdx-1 is produced early in rodent pancreatic progenitors, and is essential to acinar, ductal, and islet endocrine cell formation [5]. Both homozygous and heterozygous mutations in the PDX-1 gene have been identified in humans, which lead, respectively, to complete agenesis of the pancreas and a form of maturity-onset diabetes in the young known as MODY 4 [5].

NeuroD1/Beta2 is a basic helix-loop-helix (bHLH) transcription factor, which binds at the conserved insulin E1 (-100/-91 bp) site in a complex with ubiquitously expressed E-box proteins. NeuroD1-null mice die of severe diabetes shortly after birth due to its crucial role in β-cell formation [19]. Moreover, deletion of NeuroD1 specifically in adult β cells in vivo causes glucose intolerance and loss of expression of many genes associated with cell maturation and function [20]. Mutations in human NEUROD1 predisposes one to another form of maturity-onset diabetes in the young, MODY 6 [21], presumably because of its importance in the production and maintenance of fully functional glucose-responsive β cells.

The MafA activator is a basic leucine zipper protein, which binds as a dimer to the conserved insulin C1/RIPE3b1 (-118/-107 bp) element. MafA, and the only additional islet synthesized large Maf family member, MafB, are expressed unusually late in pancreatic cell development in relation to other islet-enriched transcription factors. In rodents, MafB is principally present in developing α cells and β cells, and then becomes restricted to α cells soon after birth [22,23]. In contrast, MafA is only found in β cells, with expression first detected during the principal wave of insulin-positive cell production at embryonic day 13.5 in mice [24]. In human islets, MafB is not
only present in α cells but also co-produced with MafA in β cells [25]. A novel role for MafA and MafB in β-cell maturation and function was revealed upon comparing their properties to other islet-enriched transcription factor mutant mice. While islet cell populations are either lost or re-specified in most transcription factor knockout mice [5], the principal defect in MafB" embryonic mice was reduced insulin and glucagon hormone expression, with no change in endocrine cell numbers or islet cell identity [22,23]. In contrast, islet cell development was unaffected in MafA[+/-] [26] and MafA[ΔPanc] [27] mice, although glucose-regulated insulin secretion was compromised in adults.

Overall, a highly sophisticated network of transcription factors provides the infrastructure for precisely regulating insulin gene transcription, which relies on the cooperative and synergistic interactions between transcription factors and recruited coactivators. Significantly, these factors are also crucial in regulating many other β-cell genes, resulting in a variety of distinct developmental and adult phenotypes in studies of total and conditional transcription factor knockout mice. In the following section, the effect of increasing glucose levels, the most physiologically impactful mediator of β-cell function on Pdx-1, NeuroD1, and MafA will be presented (Figure 6.1).

**Physiologic regulation of insulin gene expression**

**Glucose regulation of insulin gene transcription factors**
Pdx-1 is mainly localized at the nuclear periphery at low glucose concentrations (1-2 mM) in β cells, and becomes phosphorylated and shuttles into the nucleus in response to concentrations stimulating insulin secretion. Many different signaling pathways have been shown to regulate the nucleo-cytoplasmic shuttling and transactivation potential of Pdx-1 under these conditions, including glycogen synthase kinase 3 (GSK3), p38/stress-activated protein kinase, phosphatidylinositol 3-kinase (PI3K), atypical protein kinase C isoforms, mitogen-activated protein kinase (MAPK), and Per-Arnt-Sim (PAS) kinase [6]. Glucose also appears to regulate the interaction of Pdx-1 with various transcriptional coregulators. Thus, Pdx-1 is associated with the histone deacetylases HDAC-1 and HDAC-2 to downregulate insulin gene expression under low, non-insulin stimulating glucose concentrations [28]. However, the HDAC-1/2 interaction is prevented at elevated glucose levels, which promotes association with the histone acetyl-transferase p300, hyperacetylation of histone H4, and insulin gene transcription [29]. In addition, SUMOylation of Pdx-1 increases its nuclear localization as well as its protein stability and is correlated with an increase in insulin promoter activity [30]. Finally, Pdx-1 contains at least two sites of O-GlcNAcylation that increase its DNA binding activity [31,32].

In addition, glucose-induced phosphorylation of NeuroD1 regulates its nuclear localization and transactivation. However, precisely how the regulation is imposed is unknown. Treatment of pancreatic β cells with the MAPK/ERK kinase inhibitor PD98059 blocks nuclear NeuroD1 translocation, conditions that also impact Pdx-1 phosphorylation and insulin gene transcription [33]. In addition, high glucose levels induce O-GlcNAcylation of NeuroD1 [34], which appears to be important for its translocation into the nucleus. Exactly how these distinct modifications control NeuroD1 activity in islet β cells in vivo still needs to be explored.

Increased MafA protein and DNA activity regulates glucose-dependent insulin gene transcription, with phosphorylation modulating insulin enhancer binding [35]. In addition, the transactivation potential of MafA is potentiated by GSK3-mediated phosphorylation within the N-terminal region which allows recruitment of the P/CAF coactivator [36]. However, precisely how phosphorylation regulates these properties of MafA remains to be determined.

**Glucose regulation of insulin mRNA stability**
In addition to its major effects on transcription of the insulin promoter, glucose markedly stabilizes preproinsulin mRNA. Indeed, the half-life of the message has been estimated to increase from 29 hours to 77 hours when switched from low to high glucose. Two elements located in the 5'-untranslated region of the mRNA molecule have been proposed as mediators of the glucose-stabilizing effect, the conserved UUGAA sequence and the pyrimidine-rich sequence (insPRS) [37]. Stabilization appears to involve binding of a polypyrimidine tract-binding (PTB) protein to the insPRS, as binding was induced by glucose and prevented upon mutating the core PTB binding site.

**Glucose regulation of the insulin gene in vivo**
Variations in the amount of insulin mRNA at any given time represent the net effect of metabolic, hormonal, and neuronal stimuli on insulin gene transcription and mRNA stability. From the in vitro effects described earlier, it is predicted that increases in blood glucose should rapidly elevate preproinsulin mRNA levels in the endocrine pancreas (due to rapid stimulation of transcription), whereas a decrease in blood glucose would be followed by a slow decline in preproinsulin mRNA levels (due to reduced transcription and the long half-life of the message). Indeed, early studies in rats showed that 4 days of starvation are necessary to detect significant decreases in preproinsulin mRNA, which returned to basal values only 6 hours after re-feeding [38] or 12 hours after glucose injection [39]. The delay in the disappearance of the message upon fasting appears to be directly due to the stability of the mRNA, as insulin-induced hypoglycemia is followed by a rapid (2-hour) decrease in the level of precursors for insulin mRNA [40].

An additional level of complexity in the regulation of insulin gene expression in vivo is the potential interaction between β cells and non-β (i.e., α, δ, and PP) cells within the intact islet. Indeed, it is known that individual β cells behave differently from one another in terms of insulin secretion than in the context of the islet [41], a phenomenon which has also been demonstrated at the level of insulin gene transcription [42]. Thus, the global response measured in an entire islet represents the integration of
the individual responses. This fact should be kept in mind when interpreting experiments performed in clonal cell lines, which lack the level of regulation provided by islet architecture and the neighboring non-β cells.

**Regulation of insulin gene expression by glucagon-like peptide-1 (GLP-1)**

GLP-1 is an incretin hormone which strongly potentiates glucose-induced insulin secretion and is the target of a number of type 2 diabetes drugs [43]. Most of the biologic effects of GLP-1 are mediated by its G protein-coupled receptor expressed on the β-cell surface [44]. GLP-1 stimulates insulin gene expression by various mechanisms. First, it directly activates the cyclic-AMP (cAMP) response element (CRE) within the 5′-proximal control sequences of the insulin gene, by a mechanism which seems at least partly independent from protein kinase A activation [45]. Second, it can augment the glucose-stimulated binding activity of Pdx-1 [46]. Third, it can stimulate transcription of the PDX-1 gene (the promoter of which also contains a CRE) [47]. Fourth, GLP-1 potentiates glucose-induced insulin gene transcription by activating NFAT (of which there are three binding sites on the rat insulin 1 promoter) via calcium/calmodulin-dependent protein phosphatase 2B (calcineurin) activation in response to a rise in intracellular calcium [48].

**Physiologic inhibitors of insulin gene expression**

Several physiologic inhibitors of insulin secretion also impair expression. Epinephrine and somatostatin, two hormones acting through G protein-coupled receptors and known inhibitors of insulin release [49], also inhibit the rate of insulin gene transcription [50]. In addition, somatostatin decreases the stability of insulin mRNA [51]. Glucagon, a key hormone in the counterregulatory response to hypoglycemia in vivo, stimulates expression of the inducible cAMP early repressor, which inhibits insulin gene transcription [52]. Finally, the adipocyte-secreted hormone leptin decreases insulin gene expression [53]. Thus, physiologic modulators of insulin secretion coordinately inhibit insulin gene expression, thereby ensuring that the long-term biosynthetic rate of insulin matches the secretory demand.

**Short-term regulation of insulin gene expression by insulin**

Whether insulin has a functionally relevant role in the regulation of insulin transcription and biosynthesis remains a debated issue [54,55]. Based on the observation that a rapid (within minutes) transcriptional response to glucose in insulin-secreting cells was mimicked by depolarizing agents and exogenous insulin and was suppressed by inhibiting insulin release or the PI3 kinase pathway, it was proposed that insulin acts in an autocrine manner to stimulate insulin gene transcription by binding to the insulin receptor [54]. However, subsequent studies have failed to detect significant changes in preproinsulin mRNA levels upon short-term exposure to glucose and the physiologic relevance of an autocrine positive feedback on the β cell has been questioned [55].

**Dysregulation of the insulin gene**

There is convincing evidence that abnormalities in insulin gene sequence or function play a role in pancreatic β-cell dysfunction in type 2 diabetes. Abnormalities in the insulin gene structure consist of rare control region mutations, while insulin gene expression appears to be reduced by metabolic conditions associated with the diabetic state.

**Polymorphisms of the insulin gene**

The diabetes susceptibility gene IDDM2 has been mapped to the insulin variable number of tandem repeats (VNTR), a highly polymorphic region located 360 bp upstream of the transcription initiation site in the human insulin gene [56]. VNTRs are classified as class I, II, or III, depending on the number of tandem repeats. Whereas the short class I VNTR gene predisposes to type 1 diabetes, the long class III allele is protective. Precisely how VNTR polymorphisms confer susceptibility to or protection from diabetes remains uncertain, although recent evidence clearly suggest that the VNTR determines expression levels of insulin in the thymus and, in turn, the numbers of insulin-specific autoreactive T cells [56].

**Glucotoxicity and the insulin gene**

The glucotoxicity hypothesis proposes that chronic hyperglycemia is deleterious to β-cell function by contributing to the deterioration of insulin secretion [57]. These adverse effects of chronically elevated glucose levels include, but are not limited to, impairment of insulin gene expression in insulin-secreting cells, isolated rat and human islets, and animal models of diabetes [58]. The molecular mechanisms underlying glucotoxicity at the insulin gene involve decreased expression of Pdx-1 and MafA (Figure 6.2), as well as increased expression of C/EBPβ which directly binds the NeuroD1/Beta2, thereby preventing formation of the NeuroD1/Beta2:E47 activator complex required for insulin E1 stimulation [59]. In addition, binding of a C/EBPβ-NFAT complex at the A2C1 element of the insulin promoter under glucotoxicity prevents the formation of the MafA-NFAT complex at that site required for normal glucose stimulation of insulin transcription [60].

The biochemical mechanisms whereby chronically elevated glucose impairs insulin gene expression have received considerable attention in the past few years. The prevailing hypothesis is that high glucose induces the excessive production of reactive oxygen species (ROS) and the formation of advanced glycation end-products (AGE) [61,62]. This hypothesis is supported by in vivo observations. For example, treatment of Zucker diabetic fatty (ZDF) rats with the antioxidant N-acetylcysteine normalizes plasma glucose levels and restores insulin secretion, insulin content, and preproinsulin mRNA levels [63]. Similarly, overexpression of glutathione peroxidase-1 in db/db mice reversed hyperglycemia and restored MafA nuclear localization [64].
Figure 6.2 Molecular mechanisms of action for glucotoxicity and glucolipotoxicity at the level of insulin gene expression. Under physiologic conditions, MafA and Pdx-1 are two critically important regulators of the insulin promoter, and respectively bind to the C elements and the A boxes (upper panel). Glucotoxicity greatly diminishes protein levels of Pdx-1 and MafA, the former through a posttranscriptional mechanism and the latter through a posttranslational mechanism. These abnormalities lead to decreased insulin mRNA, insulin content, and glucose-induced insulin secretion, and are reversible only in the early stages of glucose toxicity (middle panel). Under glucolipotoxic conditions, MafA expression is inhibited, whereas Pdx-1 is affected at the post translational level in its ability to translocate to the nucleus. This results in decreased insulin gene expression (lower panel). Source: Poitout 2008 [58]. Reproduced with permission of The Endocrine Society.

Oxidative stress-mediated impairment in Pdx-1 binding activity is prevented by overexpression of a dominant-negative c-jun N-terminal kinase (JNK), and is mimicked by overexpression of wild-type JNK [65]. In addition, chronic exposure to elevated glucose levels may lead to dedifferentiation with loss of genes associated with β-cell function and overexpression of genes normally repressed in differentiated β cells [66,67]. For instance, the c-myc transcription factor is upregulated in diabetic islets [68] and is induced by high glucose in normal islets [69]. In turn, c-myc can inhibit insulin gene transcription [70] by competing for NeuroD1/Beta2 binding at the E-box [71].

Endoplasmic reticulum (ER) stress has also been proposed to contribute to the mechanisms of glucotoxicity independently from oxidative stress [72]. However, alleviation of ER stress by chemical chaperones in glucose-cultured islets improves insulin secretion but not intracellular insulin content, suggesting that ER stress may be involved in defective insulin secretion but not impaired insulin biosynthesis under glucotoxic conditions [73].

**Glucolipotoxicity and the insulin gene**

Like chronic hyperglycemia, hyperlipidemia has been proposed to contribute to β-cell dysfunction in type 2 diabetes [74]. Most of the deleterious effects of chronically elevated lipid levels on the β cell require the concomitant presence of hyperglycemia, a phenomenon referred to as glucolipotoxicity [75]. Amongst its many functional consequences, glucolipotoxicity impairs insulin gene expression via a transcriptional mechanism that involves de novo synthesis of ceramide and defective function of Pdx-1 and MafA [76,77]. Importantly, defective Pdx-1 function and insulin gene expression are also observed in an in vivo model of glucolipotoxicity in rats following a 72-h infusion of glucose and Intralipid, a lipid emulsion which raises circulating
fatty acid levels when co-infused with heparin [78,79]. It is interesting that glucotoxicity and glucolipotoxicity both affect Pdx-1 and MafA function, albeit by different mechanisms: glucotoxicity alters Pdx-1 mRNA expression [80] and MafA nuclear localization [64], while in glucolipotoxicity Pdx-1 is retained in the cytosolic compartment while MafA mRNA expression is reduced [77] (Figure 6.2).

How de novo ceramide synthesis from palmitate, in the presence of elevated glucose, alters the function of Pdx-1 and MafA and leads to defective insulin gene expression remains unknown. One possible candidate is the serine/threonine kinase PAS kinase, which regulates glucose-induced insulin gene transcription [81]. In insulin-secreting cells and isolated islets, we observed that overexpression of PAS kinase protects from the negative effects of palmitate on the insulin gene [82]. Recent data suggest that this could be mediated by PAS kinase inactivation of GSK3β (via phosphorylation at Ser9) and alleviation of GSK3β-mediated serine phosphorylation of Pdx-1 and proteasomal degradation ([83] and M. Semache, G. Fontés, S. Fogarty, C. Kikani, M. B. Chawki, J. Rutter and V. Poitout, unpublished data). A second candidate mediator of ceramide inhibition of the insulin gene is c-Jun N-terminal kinase (JNK). In support of this possibility, palmitate was shown to activate JNK in β cells which results in a decrease in insulin gene transcription [84].

Relevance to human type 2 diabetes
Recent studies in human islets support the notion that defective insulin gene expression may play a role in human type 2 diabetes. First, in islets isolated from pancreata of 13 type 2 diabetic cadaveric organ donors high levels of oxidative stress markers as well as low levels of glucose-induced insulin secretion, reduced insulin mRNA, but increased levels of Pdx-1 and FOXO1 mRNAs have been observed [85]. Second, nuclear expression of MafA is decreased in human diabetic islets [86]. Third, DNA methylation of the insulin promoter is increased in type 2 diabetic patients and correlates negatively with insulin gene expression and positively with hemoglobin A1c levels [87].

Insulin biosynthesis
Introduction
The previous section outlines that insulin gene transcription is a highly controlled process. The product of this process, proinsulin mRNA, is unusually stable in pancreatic β cells and it is further stabilized as glucose concentrations increase [37]. As such, there is normally an abundant source of proinsulin mRNA in the β-cell cytosol available for translation. Actually, it is the specific regulation of proinsulin mRNA translation that is the predominant control mechanism for insulin production in the β cell under normal circumstances. This enables the β cell to rapidly replenish insulin stores back to optimal levels, after they have been depleted by stimulated insulin secretion, and is more economic energy-wise to the β cell, since translational control of insulin production bypasses the need for insulin gene transcription and proinsulin mRNA maturation.

Structure of the insulin molecule
The primary structure A- and B-chain of insulin itself has been known for close to 50 years [88]. However, it was not until at least 10 years after this discovery that it was realized insulin is actually synthesized as a single polypeptide chain precursor molecule, proinsulin (Figure 6.3) [88]. The N-terminal signal peptide (24 amino acids) is cleaved cotranslationally to yield proinsulin. The proinsulin molecule is a 12-kDa single chain polypeptide that encompasses the B-chain (30 amino acids) and the A-chain (21 amino acids) of insulin joined by the connecting peptide, C-peptide (Figure 6.3). Proinsulin to insulin conversion occurs by cleavage at two dibasic amino acids sequences by the B-chain/C-peptide and C-peptide/A-chain junctions to release the C-peptide moiety yielding the insulin molecule with the two independent disulphide-linked A- and B-chains correctly aligned [88].

Proinsulin biosynthesis: translation and translocation
Essentially, translation of proinsulin mRNA to proinsulin protein occurs in a fashion typical of most eukaryotic mRNAs destined to enter the cell’s secretory pathway [88–90]. During the translation process, the emerging signal sequence of proinsulin binds the signal recognition particle (SRP) that then docks to the SRP-receptor, which is an integral ER membrane protein. This locates the proinsulin mRNA/ribosomal translational complex to the ER, which is the major site of proinsulin biosynthesis in the β cell. As SRP binds to the SRP receptor, the nascent signal peptide of the newly forming proinsulin dissociates and is transferred to another ER integral membrane protein, the signal sequence receptor (SSR). SSR is part of a “translocation pore” that facilitates transport of the newly forming proinsulin polypeptide across the ER membrane into the ER lumen, marking the entrance of the newly synthesized preproinsulin into the β-cell’s secretory pathway. The signal peptide is cleaved by another RER protein, the signal peptidase, resulting in the nascent proinsulin molecule located to the ER lumen. There, the proinsulin molecule undergoes appropriate folding, assisted by the molecular chaperons and formation of disulfide bonds catalyzed by ER disulfide isomerase activity [88].

Proinsulin biosynthesis: effectors and stimulus-response coupling mechanisms
Proinsulin biosynthesis is translationally controlled by certain nutrients, neurotransmitters, and hormones, but glucose is the most physiologically relevant [91]. This translational control response to glucose is rapid. Significant glucose-induced proinsulin biosynthesis can be observed after a 20–30-min lag period that reaches a maximum rate (~20–30-fold increase above basal) by 60 min [91]. Of the peptide hormones that stimulate
proinsulin biosynthesis, which results in the increased glucose metabolism in the β cell, resides in the 5′- and 3′-untranslated regions (UTRs) of preproinsulin mRNA [92]. In the 3′-UTR of preproinsulin mRNA, just downstream of the polyadenylation signal, there is a highly conserved primary sequence containing a UUGAA cis-element core, that has been reported to be involved in glucose-regulated preproinsulin mRNA stability in addition to the pyrimidine-rich sequence (insPRS) [37,92] (see earlier). There is some degree of cooperativity between the preproinsulin mRNA 5′- and 3′-UTRs for the specific glucose-induced translational control of proinsulin biosynthesis, but it appears that the 5′-UTR preproinsulin mRNA has the major influence [92]. There is also a conserved cis-element that is required for glucose-induced translational control of proinsulin biosynthesis, named ppIGE (for preproinsulin glucose element) [92]. The ppIGE has a highly conserved ppIGE palindromic core of GUCxₘ CUG or GUₙxₘ UUG (where n ≤ 4 bases). A cytosolic protein trans-acting factor

Proinsulin biosynthesis: translational control mechanism

Glucose modestly increases general protein synthesis in the β cell ~1.5–2-fold. However, the effect of glucose on proinsulin synthesis translation is much greater, and can reach ≥10-fold stimulation above basal [91,92]. This indicates a specific effect of glucose on translational control of proinsulin biosynthesis. Specific control of glucose-induced proinsulin biosynthesis in the β cell, resides in cis-elements in the 5′- and 3′-untranslated regions (UTRs) of preproinsulin mRNA itself [92]. In the 3′-UTR of preproinsulin mRNA, just downstream of the polyadenylation signal, there is a highly conserved primary sequence containing a UUGAA cis-element core, that has been reported to be involved in glucose-regulated preproinsulin mRNA stability in addition to the pyrimidine-rich sequence (insPRS) [37,92] (see earlier). There is some degree of cooperativity between the preproinsulin mRNA 5′- and 3′-UTRs for the specific glucose-induced translational control of proinsulin biosynthesis, but it appears that the 5′-UTR preproinsulin mRNA has the major influence [92]. There is also a conserved cis-element that is required for glucose-induced translational control of proinsulin biosynthesis, named ppIGE (for preproinsulin glucose element) [92]. The ppIGE has a highly conserved ppIGE palindromic core of GUCₙₓₘ CUG or GUₙₓₘ UUG (where n ≤ 4 bases). A cytosolic protein trans-acting factor

**Figure 6.3** Primary structure of human preproinsulin indicating the N-terminal signal peptide, followed by the insulin B-chain, the connecting-peptide (C-peptide), and then the C-terminal insulin A-chain. The two dibasic proteolytic conversion sites and the endoproteases that cleave at these are also indicated (PC-3 at Arg³¹, Arg³²; and PC-2 at Lyv⁶⁴, Arg⁶⁵). The signal peptide + insulin + C-peptide moieties comprise preproinsulin, and the insulin + C-peptide moieties comprise proinsulin.
(ppIE-BP) binds to this translational control ppIGE cis-element of preproinsulin mRNA in a glucose-dependent manner, but the identity of the ppIGE-BP has yet to be revealed [92]. It should be noted that proinsulin is only one of a small subset of β-cell proteins (~50 in all) [92] whose biosynthesis is regulated by glucose at the translational level. These are mostly β-granule proteins, including the proinsulin processing endopeptidases, PC2 and PC1/3 [92]. Indeed, the ppIGE is also conserved in the 5′-UTR of the majority insulin secretory granule proteins' mRNAs, and thus, the glucose-induced specific translational control of proinsulin biosynthesis and that for insulin secretory granules can be coordinated and is the principal control mechanism for insulin secretory biogenesis in β cells [92]. The glucose-induced translational control of the proinsulin processing endopeptidases, proPC2 and proPC3, provides a means whereby proinsulin conversion is not compromised upon increased proinsulin biosynthesis [92].

### Transport of proinsulin from the ER to Golgi apparatus

After proinsulin is translocated into the lumen of the ER, it is then delivered in transport “COPI-coated vesicles” to the cis-Golgi apparatus [93] (Figure 6.4). Up until relatively recently, it was thought that newly synthesized proinsulin was passed from the cis-Golgi network “stack” via the medial- to the trans-stack of the Golgi apparatus stacks in “COPI-coated vesicles, but a landmark study conducted in β cells using electron microscope tomography showed that the Golgi apparatus is actually one continuous organelar compartment, and not a series of stacks [94]. As such, newly synthesized proinsulin traverses through the lumen of the β cell's Golgi apparatus to the trans-Golgi network (TGN) [94], where it accumulates in clathrin-coated regions [93]. This is the site of secretory granule biogenesis (Figure 6.4). The means by which newly synthesized proinsulin (and other select proteins destined to the β granule), is specifically targeted to sites of β-granule biogenesis in the TGN remains a matter of debate [95]. However, it is known to be a highly efficient process, with >99% efficient of newly synthesized proinsulin sorted to the β granule and regulated secretory pathway under normal conditions [91,95].

Generally, analogous to other neuroendocrine cells, the process of β-granule biogenesis should also require other factors including intraluminal acidic pH 6.5, Ca²⁺, ATP, GTP-hydrolysis cytosolic proteins and perhaps protein tyrosine phosphorylation [88,91,93,95]. Although β-granule biogenesis occurs in limited clathrin-coated regions of the TGN [93], the role that clathrin itself plays is unclear although is likely involved in the process of a newly formed immature β granule “budding off” the TGN. An immature β granule then undergoes a maturation process [93,95]. Maturation of β granules involves proinsulin conversion, progressive intragranular acidification, loss of the clathrin-coated regions, and formation of hexameric insulin crystals [91,93]. Acidification provides the correct intragranular pH (pH 5.0–5.5) for proinsulin processing to proceed [91], and optimal insulin crystal formation around insulin's isoelectric point (pKi 5.3) [88]. Delivery of newly synthesized proinsulin to an immature β granule occurs around 30–40 min posttranslational where proinsulin processing begins and is >90% completed ~3 h later [91,93] (Figure 6.4).

### Proteolytic enzymes of proinsulin conversion

The major site for processing of proinsulin to biologically active insulin is the immature secretory granule compartment of the β cell [91,93] (Figure 6.4). Production of insulin (and C-peptide) occurs via limited proteolysis of the proinsulin precursor molecule, which is catalyzed by two Ca²⁺-dependent endopeptidases, PC2 and PC1/3 and a Ni²⁺-dependent exopeptidase, CP-H [88,91]. There are two dibasic sites on the human proinsulin molecule: Arg₃¹, Arg₃² and Lys₆⁴,Arg₆⁵, that signal limited endoproteolytic cleavage of proinsulin to excise the C-peptide moiety and to generate insulin with its disulphide-linked A- and B-chains correctly aligned (Figure 6.3). Endoproteolytic peptide bond cleavage of proinsulin occurs on the carboxylic side of the Arg₃¹, Arg₃² or Lys₆⁴,Arg₆⁵, followed by rapid and specific exopeptidic removal of the newly exposed basic amino acids by CP-H [88,91]. The two distinct β-granule proinsulin-processing endopeptidase activities were originally discovered as Ca²⁺-dependent with an acidic pH optimum and were later identified as the PC1/3 and PC2 endopeptidase genes [88,91].

A scheme of proinsulin conversion is illustrated in Figure 6.5. Proinsulin conversion could occur by two possible routes. Either PC2 first cleaves on the carboxylic side of Lys₆⁴-Arg₆⁵ to yield a split 65,66 proinsulin intermediate, followed by CP-H trimming of the newly exposed lysine and arginine residues to yield des 64,65 proinsulin. Then PC1/3 can then cleave des 64,65 proinsulin to Arg₃¹, Arg₃², which together with CP-H trimming of the exposed arginine residues, yields insulin and C-peptide (Figure 6.5). Alternatively, PC1/3 first could cleave at the carboxylic side of Arg₃¹, Arg₃² to yield a split 32,33 proinsulin intermediate, followed by CP-H trimming of the revealed arginine residues to yield des 31,32 proinsulin. PC2 can then cleave des 32,33 proinsulin at Lys₆⁴-Arg₆⁵, which together with CP-H trimming of the lysine and arginine residues, yields insulin and C-peptide (Figure 6.5). However, PC2 has a much stronger preference for the des 31,32 proinsulin substrate than proinsulin, whereas PC1/3 has an equivalent preference for proinsulin or des 64,65 proinsulin substrates [88,91]. As such, in humans, the sequential processing of proinsulin via des 31,32 proinsulin is the predominant route, where PC1/3 cleaves intact proinsulin first, followed by PC2 cleavage of des 31,32 proinsulin (Figure 6.5). This is consistent with the presence of the des 31,32 proinsulin conversion intermediates in the human circulation but negligible levels of des 64,65 proinsulin [91].

PC2, PC1/3 and CP-H are expressed in most neuroendocrine cells where they are involved in posttranslational processing of other prohormone precursors [88]. The role of these proteolytic enzymes in proinsulin conversion has been substantiated in
Figure 6.4 Cell biology of proinsulin trafficking and processing. The site of preproinsulin biosynthesis is on the ribosomes of the ER. The signal peptide is cleaved co-translationally enabling proinsulin translocation into the ER lumen. The newly synthesized proinsulin is then transported to the cis-Golgi and transported though the stacks of the medial- and trans-Golgi to clathrin-coated regions of the trans-Golgi network (TGN). Immature β granules bud off the TGN which is the major site of proinsulin conversion (~30 min posttranslationally). Proinsulin conversion proceeds as a β granule matures. Mature granules form the intracellular insulin storage compartment of the β cells and do not undergo Ca^{2+}-dependent exocytosis unless triggered by an appropriate stimulus. The left panel indicates the intracellular compartments in which proinsulin is sequentially transported through, and the right panel indicates the kinetics of the preproinsulin biosynthetic/processing/secretory process in these compartments.

various gene-deletion studies. PC2, PC1/3 or CP-H deficiencies render multiple endocrine deficiencies. PC2 knockout mice have defective proinsulin processing with increased levels of the split proinsulin conversion intermediate des 31,32 proinsulin, consistent with PC2 preferentially cleaving at the Lys^{64}, Arg^{65} site on proinsulin, and the preferred sequential proinsulin processing route [88] (Figure 6.3). Indeed, PC2 null mice have a polyendocrine phenotype the most obvious being fasting hypoglycemia and glucose intolerance due to a deficiency of circulating glucagon levels rather than increased insulin levels [88]. A human mutation of both PC1/3 alleles exists, which results in negligible PC1/3 activity [88]. This generates a complicated phenotype of multiple endocrine disorders due to general abnormal prohormone processing, one of which is very low insulin levels and high proinsulin levels, together with abnormal glucose homeostasis, consistent with defective proinsulin processing [88]. A very similar phenotype is found in the PC1/3 null transgenic mouse model [88]. Finally, the obese Fat/Fat mice have been found to have a mutation in the PC-H gene resulting in negligible CP-H activity [91]. These CP-H null animals are hyperproinsulinemic, suggesting that CP-H trimming off of basic amino acids after PC2 and
Insulin gene expression and biosynthesis

Insulin

Figure 6.5 Enzymatic proteolytic conversion of proinsulin. There are two potential pathways of proteolytic conversion of proinsulin to insulin. Either PC2 first cleaves proinsulin at Lys\(^{64}\)-Arg\(^{65}\) to yield split 65,66 proinsulin, followed by CP-H trimming of the newly exposed lysine and arginine residues to yield des 64,65 proinsulin. PC3 can then cleave des 64,65 proinsulin at Arg\(^{31}\), Arg\(^{32}\), which together with CP-H trimming of the exposed arginine residues, yields insulin and C-peptide. Alternatively, PC3 cleaves at Arg\(^{31}\), Arg\(^{32}\) to yield a split 32,33 proinsulin, followed by CP-H trimming of the revealed arginine residues yields des 31,32 proinsulin. PC2 can then cleave des 32,33 proinsulin at Lys\(^{64}\)-Arg\(^{65}\), which together with CP-H trimming of the lysine and arginine residues, yields insulin and C-peptide. In human β cells, the route via des 31,32 proinsulin predominates as illustrated by the larger size of this pathway.

PC1/3 endopeptidic cleavage accelerates proinsulin proteolytic maturation through to insulin [91]. Moreover di-arginyl insulin (that has >50% reduced biological activity) rather than insulin is produced indicating the role that CP-H plays in trimming basic amino acids during the proinsulin conversion process [91].

Regulation of proinsulin conversion

PC2 and PC1/3 are Ca\(^{2+}\)-dependent enzyme activities with an acidic pH 5–5.5 optimum [91]. Fortunately, the β granule contains an intraorganellar environment of 1–10 mM free Ca\(^{2+}\) and acidic pH 5.5, which ideally suits the requirements for optimal PC2, PC1/3 and CP-H activities within this organelle. This also ensures that insulin is produced mainly in the intracellular β-granule compartment in which it is stored [91]. To render PC2 and PC1/3 fully active for proinsulin processing in a newly formed β granule, it follows that activation of the proton-pumping ATPase and Ca\(^{2+}\)-translocation proteins [91] are key regulatory events to control proinsulin conversion.

Both PC2 and PC1/3 are initially synthesized as preproprotein precursor molecules themselves, with the “pre” signal peptide region enabling translocation into the RER lumen during translation as with the signal peptide region of proinsulin (see earlier). ProPC2 and proPC1/3 are transported to β granules along with their proinsulin substrate in the β cell’s secretory pathway and undergo maturation beginning in the TGN [91]. However, unlike proinsulin, proPC2 and proPC1/3 are thought to be accompanied by individual chaperon molecules, 7B2 and proSAAS, respectively, that specifically inhibit these endopeptidases’ activity [91]. Proteolytic cleavage of 7B2 by PC2 in the TGN/immature β-granule compartment alleviates the inhibition on proPC2 promoting its maturation and activation to mature PC2. Indeed, 7B2 has an important role in controlling PC2 activity in vivo. The 7B2 knockout mouse has multiple neuroendocrine disorders, similar yet more severe than the PC2 null mouse [91]. In contrast, the role of proSAAS in regulating proinsulin processing is doubtful, since proSAAS null mice have normal insulin production and proSAAS is not highly expressed in β cells [96,97].

As previously indicated, the biosynthesis of proPC2 and proPC1/3 is stimulated predominately at a translational level coordinately with that of proinsulin [88,91,92]. In the long term (>12 h), glucose also regulates PC2 and PC1/3 gene transcription in parallel with the preproinsulin gene [91]. Thus, it seems that proinsulin conversion is adaptable to changes in glucose by coordinate regulation of the endopeptidases that catalyze processing [91,92].
The mature β-granule storage pool
A mature β granule is retained from anywhere between a few hours to several days, awaiting transport to the β cell’s plasma membrane and exocytosis under stimulatory conditions, characteristic of a regulated secretory pathway [88,92] (Figure 6.4). It should be noted that under normal conditions, the storage compartment of insulin in mature β granules far exceeds the compartment undergoing transport/exocytosis, so that during a 1-h stimulation by glucose only ~1–2% of the insulin content of a primary islet β cell is secreted [92]. The insulin content of a β cell is kept at a relatively constant level under normal physiologic conditions where secreted insulin is rapidly replaced at the biosynthetic level. However, in the long term there is also an additional regulatory component that maintains insulin stores at optimal levels, via insulin degradation [92]. The half-life of a β granule is several days, but if it is not used for exocytosis it is eventually degraded by fusion with lysosomal compartments by autophagy (also known previously as crinophagy) [92].

Dysfunctional proinsulin processing in diabetes
In type 2 diabetes where there is hyperinsulinemia to compensate for peripheral insulin resistance, an increased proportion of the secreted insulin is actually proinsulin or split proinsulin conversion intermediates (mostly des 31,32 proinsulin) so that it is also a hyperproinsulinemic state [91]. It is possible that genetic defects in the proinsulin conversion enzyme genes or the insulin gene itself hamper proinsulin conversion, resulting in an increased proportion of proinsulin secreted. However, such genetic mutations are very rare, yet hyperproinsulinemia is a common trait of type 2 diabetes [91]. As such, an increased proportion of secreted proinsulin likely occurs as a consequence of β-cell secretory dysfunction in type 2 diabetes [91].

In common obesity-linked type 2 diabetes there is chronic hyperglycemia and dyslipidemia [88,91,92]. As a consequence, the β cell is working very hard, with both proinsulin synthesis and insulin secretion are upregulated in an attempt to compensate for peripheral insulin resistance. Normally in β cells there is preferential exocytosis of newly formed β granules, but under such chronic stimulation from hyperglycemia/hyperlipidemia newly synthesized proinsulin is not retained long enough to be fully converted to insulin and C-peptide, and as a consequence a greater proportion of proinsulin (as well as des 31,32 proinsulin) is secreted [88,91]. It should also be noted that chronic dyslipidemia adversely affects secretory capacity of β cells. Elevated fatty acid levels increase the amount of insulin secreted from the β cell, but in contrast, fatty acids modestly inhibit glucose-induced proinsulin biosynthesis, which in turn markedly decreases insulin content of islet β cells in vivo [92]. A similar situation might also be envisaged with the prolonged use of sulfonlyureas, which though potent inducers of insulin secretion, do not stimulate proinsulin synthesis and decrease insulin content [92], thus also reducing the insulin secretory capacity of the β cell. In general, the chronic hyperglycemia and dyslipidemia in obesity-linked type 2 diabetes are constantly making the β cell work harder to produce sufficient insulin to compensate for increased metabolic load and peripheral insulin resistance [91,92]. But in the long run this eventually leads to β-cell dysfunction of which the hyperproinsulinemia is symptomatic. Interestingly, if the β cell in type 2 diabetes patients is allowed to rest, the β-cell secretory dysfunction in vivo is reduced. This emphasizes the importance of protecting β-cell mass and function in the treatment of obesity-linked type 2 diabetes [88,91,92].

Acknowledgments
Work from our laboratories cited in this chapter was supported by the following grants: R01 DK-050610, R01 DK-055267, and the JDRF and Brehm Coalition (C.J. Rhodes); R01 DK-058096 and the Canada Research Chair in Diabetes and Pancreatic Beta-cell Function (V. Poitout), R01 DK 50203, R01 DK-055091 and R01 DK-042502 (R. Stein).

References


31 Gao Y, Miyazaki J, Hart GW: The transcription factor PDX-1 is post-translationally modified by O-linked N-acetylglucosamine and this modification is correlated with its DNA binding activity and insulin secretion in min6 beta-cells. *Archives of Biochemistry and Biophysics* 2003;415(2):155–163.


Insulin gene expression and biosynthesis


CHAPTER 7

β-Cell biology of insulin secretion

Susumu Seino1, Tadao Shibasaki2 and Kohtaro Minami2

1Division of Molecular and Metabolic Medicine, Department of Physiology and Cell Biology, Kobe University Graduate School of Medicine, Kobe, Japan
2Division of Cellular and Molecular Medicine, Department of Physiology and Cell Biology, Kobe University Graduate School of Medicine, Kobe, Japan

Key points

• Pancreatic β cells secrete insulin in response to glucose.
• Glucose-induced insulin secretion (GIIS) requires metabolism of glucose.
• Glucose generates both triggering and amplifying signals.
• GIIS is modulated by hormones, neurotransmitters, and nutrients.
• GIIS is biphasic.
• There are different pools of insulin granules.
• Insulin granule exocytosis depends on soluble N-ethylmaleimide-sensitive factor attachment protein receptors (SNAREs) and their associated molecules.
• G-protein-coupled receptors are involved in modulation of GIIS.
• β-Cell dysfunction is associated with the development of diabetes.

Introduction

In pancreatic β cells, glucose metabolism is essential for the regulation of insulin secretion. Glucose is taken up by glucose transporters and metabolized to generate adenosine triphosphate (ATP), which is the main driver of glucose-induced insulin secretion (GIIS). Increased cytosolic ATP causes closure of ATP-sensitive K⁺ (K_ATP) channels, depolarizing the plasma membrane, leading to the opening of voltage-dependent Ca²⁺ channels (VDCCs), which allows Ca²⁺ influx. The resultant rise in intracellular Ca²⁺ concentration ([Ca²⁺]₀) induces exocytosis of insulin granules in the triggering pathway of insulin secretion (Figure 7.1). In addition, other signals generated by glucose amplify insulin secretion. Lipid metabolism is also involved in GIIS by interacting with glucose metabolism.

Glucose induces insulin secretion in a biphasic manner: an initial component (1st phase) develops rapidly but lasts only a few minutes, and is followed by a sustained component (2nd phase). Pancreatic β cells contain at least two pools of insulin secretory granules that differ in release competence: a reserve pool (RP) that accounts for the vast majority of granules, and a readily releasable pool (RRP) that accounts for the remaining <5%. Although the prevailing hypothesis is that release of predocked granules accounts for the 1st phase and a subsequent supply of new granules mobilized for release accounts for the 2nd phase of GIIS, recent studies show that both phases involve granules that are located some distance from plasma membrane. Hormonal and neural inputs to the β cells are also important for modulating GIIS.

β-Cell metabolism

Glucose sensing glycolysis

The most prominent feature of pancreatic β cells is secretion of insulin in response to changes in the physiologic concentration of extracellular (blood) glucose, the cells possessing the capacity to sense circulating glucose levels. Glucose is transported into the β cells through facilitated glucose transporters and then is promptly phosphorylated by glucokinase in the glycolytic pathway. In rodents, the major glucose transporter is GLUT2, which is a high-capacity, low-affinity glucose transporter isoform. Although GLUT2 is the major glucose transporter in pancreatic β cells in rodents, GLUT1 is predominantly expressed in human β cells [1]. On the other hand, the glucose-phosphorylating enzyme glucokinase (hexokinase IV: a high K_m isoform of hexokinase) catalyzes the formation of glucose-6-phosphate from glucose without allosteric inhibition of the product. As glucokinase determines the rate of glycolysis, it is considered to be the molecular glucose sensor for insulin secretion in pancreatic β cells [2]. Indeed, overexpression of hexokinase shifts glucose sensitivity in a mouse pancreatic β-cell line [3] and mutations in the glucokinase gene can cause diabetes [4]. Phosphorylated glucose is then metabolized to produce pyruvate, the end product of glycolysis. As the expression of lactate dehydrogenase (LDH) is very low in pancreatic β cells [5],
Figure 7.1 Glucose-induced insulin secretion (GIIS). Glucose is transported via glucose transporters and metabolized to generate ATP. Glucokinase acts as the glucose sensor. Increased cytosolic ATP closes ATP-sensitive K⁺ (K\textsubscript{ATP}) channels, depolarizing the plasma membrane and opening the voltage-dependent Ca\textsuperscript{2+} channels (VDCCs), which allows Ca\textsuperscript{2+} influx. The rise in intracellular Ca\textsuperscript{2+} concentration ([Ca\textsuperscript{2+}]\textsubscript{i}) triggers exocytosis of the insulin granules. Glucose may generate various metabolic signals, yet to be identified, to amplify GIIS. GIIS can be modulated by hormones and neurotransmitters. VDCC, voltage-dependent Ca\textsuperscript{2+} channel; PKC, protein kinase C; PKA, protein kinase A; ATP, adenosine triphosphate; IP\textsubscript{3}, inositol 1,4,5-triphosphate; DAG, diacylglycerol.

Mitochondrial metabolism

In the mitochondrion, pyruvate is converted to acetyl-CoA by pyruvate dehydrogenase (PDH) and reacts with oxaloacetate to form citrate, an intermediate metabolite in the tricarboxylic acid (TCA) cycle. In addition, an anaplerotic pathway provides oxaloacetate for the TCA cycle directly from pyruvate by pyruvate carboxylase (PC). This reaction is also involved in the pyruvate/malate shuttle. The TCA cycle is an important metabolic circuit in terms of production of reducing equivalents in the form of NADH and FADH\textsubscript{2} for generation of ATP in the electron transport chain. Citrate is oxidized and decarboxylated to form α-ketoglutarate, which undergoes further oxidative decarboxylation to succinyl-CoA or generates glutamate by glutamate dehydrogenase (GDH). NADH is formed in these processes. Succinyl-CoA is then metabolized to succinate and converted subsequently to fumarate by succinate dehydrogenase, by which FADH\textsubscript{2} is generated. At the end of the cycle, oxaloacetate is regenerated via malate. The TCA cycle is an important aerobic pathway for the final step of the oxidation of fatty acids and certain amino acids as well as carbohydrates.

ATP generation in the respiratory chain

Activation of the TCA cycle stimulates the electron transport chain to pump H\textsuperscript{+} ions out of the mitochondrial matrix, which hyperpolarizes the inner mitochondrial membrane. The respiratory chain comprises complex I (NADH-ubiquinone reductase), II (succinate dehydrogenase), III (ubiquinol-cytochrome c reductase), and IV (cytochrome c oxidase). Complex I and II accept electrons from NADH and FADH\textsubscript{2}, respectively, and transport them to ubiquinone (coenzyme Q). Ubiquinone then transfers electrons to complex III, which is a multisubunit transmembrane protein encoded by both the mitochondrial and the nuclear genomes. Complex III transports electrons to cytochrome c, and then to complex IV, in which these electrons are transferred to oxygen (O\textsubscript{2}), producing H\textsubscript{2}O. At the same time, protons are translocated across the membrane, contributing to the proton gradient. This gradient is used by the F\textsubscript{0}F\textsubscript{1} ATP synthase complex (sometimes called complex V) to make ATP via oxidative phosphorylation. Thus, in the respiratory chain, electrons move from an electron donor (NADH and FADH\textsubscript{2}) to a terminal electron acceptor (O\textsubscript{2}) via a series of redox reactions, which are coupled to the creation of a proton gradient across the mitochondrial inner membrane. The resulting transmembrane proton gradient is used in making ATP. Synthesized ATP is translocated to cytosol by the adenine nucleotide translocator (ANT). Since ATP production is a critical signal in the triggering pathway of GIIS from pancreatic β cells (Figure 7.1), disruption of mitochondrial function causes loss of GIIS [6,7].

NADH shuttles

NADH shuttles are linked to glycolysis to generate NAD\textsuperscript{+} in supplying electrons for the respiratory chain in the mitochondria (Figure 7.2). Pancreatic β cells cannot generate NAD\textsuperscript{+} via pyruvate readily enters the mitochondrion for subsequent oxidation.
Glucose $\rightarrow$ G6P $\rightarrow$ F6P $\rightarrow$ FBP

GAP $\rightarrow$ DHAP $\rightarrow$ BPGA

Pyruvate

TCA cycle

NADH $\rightarrow$ NAD$^+$

NADH $\rightarrow$ NAD$^+$

FADH$_2$ $\rightarrow$ FAD

Mitochondrion

GP

GLU

F6P

GLU

Malate

OAA

ASP

α-KG

Aralar1

MDH

1

AST

2

AST

1

GPD

1

GPD

2

OGC

Acetyl-CoA

PDH

GLU

GDH

Lipid metabolism

In pancreatic β cells, glucose metabolism interacts with lipid metabolism, such that GIIS is associated with inhibition of fatty acid oxidation and increased lipid synthesis. The activated form of fatty acids is long-chain acyl-CoA (LC-CoA), which is generated by acyl-CoA synthetase (ACS). The fate of fatty acids is determined by malonyl-CoA, which blocks mitochondrial oxidation of fatty acids so that LC-CoA levels are increased. A high concentration of glucose activates the TCA cycle and increases anaplerotic input of OAA, which in turn elevates export of citrate from mitochondria to cytosol via the citrate-isocitrate carrier (CIC). Citrate is cleaved by ATP-citrate lyase (ACL) to OAA and acetyl-CoA, the acetyl-CoA then being carboxylated by acetyl-CoA carboxylase-1 (ACC1) to form malonyl-CoA.

Thus, glucose stimulation of β cells elevates malonyl-CoA activity of GPD2 is extremely high in pancreatic islets [5,8], and decreased activity of GPD2 or the GP shuttle may be associated with type 2 diabetes [9]. However, GPD2 deficient mice do not have impaired GIIS [10,11], indicating that the GP shuttle is not essential for GIIS in pancreatic β cells. Inhibition of the MA shuttle by aminooxyacetate in GPD2 deficient mice abolishes GIIS [10]. Thus, the two NADH shuttles complementarily operate in pancreatic β cells in regulating GIIS.

Lipid metabolism

In pancreatic β cells, glucose metabolism interacts with lipid metabolism, such that GIIS is associated with inhibition of fatty acid oxidation and increased lipid synthesis. The activated form of fatty acids is long-chain acyl-CoA (LC-CoA), which is generated by acyl-CoA synthetase (ACS). The fate of fatty acids is determined by malonyl-CoA, which blocks mitochondrial oxidation of fatty acids so that LC-CoA levels are increased. A high concentration of glucose activates the TCA cycle and increases anaplerotic input of OAA, which in turn elevates export of citrate from mitochondria to cytosol via the citrate-isocitrate carrier (CIC). Citrate is cleaved by ATP-citrate lyase (ACL) to OAA and acetyl-CoA, the acetyl-CoA then being carboxylated by acetyl-CoA carboxylase-1 (ACC1) to form malonyl-CoA.

Thus, glucose stimulation of β cells elevates malonyl-CoA...
levels [12]. Addition of LC-CoA to permeabilized β cells stimulates insulin granule exocytosis [13], suggesting a role for LC-CoA in GIIS. Inhibition of the tricarboxylic transporter in the mitochondrial outer membrane decreases GIIS but not potassium-induced insulin secretion in rat clonal β cells [14], suggesting that export of citrate from mitochondria to cytosol is involved in metabolic signaling in the regulation of GIIS.

**Metabolism-secretion coupling**

**Triggering pathway**

Glucose metabolism increases the cytosolic ATP concentration in pancreatic β cells, this rise in ATP causing closure of the K_ATP channels and depolarization of the β-cell membrane. Thus, K_ATP channels couple the cell’s metabolic state to electrical activity. The β-cell K_ATP channel is composed of two subunits: Kir6.2 as a pore-forming subunit and the sulfonylurea receptor SUR1 as a regulatory subunit. Activity of the K_ATP channel is critical for GIIS. Membrane depolarization opens VDCCs, which allows Ca^{2+} influx into β cells, the resultant rise in intracellular Ca^{2+} triggering exocytosis of insulin granules. Thus, K_ATP channels and VDCCs are major ion channels required for metabolism-secretion coupling in insulin release.

The intracellular Ca^{2+} concentration ([Ca^{2+}]_i) in pancreatic β cells is tightly regulated. Micromolar increases in [Ca^{2+}]_i are required to trigger insulin secretion [15]. Opening of the VDCCs is a common step in insulin secretion induced by glucose, sulfonlureas, and amino acids [16]. The L-type VDCC generally gives rise to a transient Ca^{2+} concentration in pancreatic β cells; modulation of VDCC activity generates changes in insulin secretion [17]. Although rapid Ca^{2+} influx through VDCCs is indispensable in GIIS, increases in [Ca^{2+}]_i can be slowly achieved by Ca^{2+} release from intracellular Ca^{2+} stores. It has been suggested that the mobilization of intracellular Ca^{2+} from ryanodine-sensitive Ca^{2+} stores by cyclic ADP-ribose generated by glucose stimulation contributes to GIIS [18], but the notion is controversial.

Among many Ca^{2+}-binding proteins that may function as Ca^{2+} sensors for vesicle fusion, the leading candidates are members of the synaptotagmin family [19]. Most of the 15 members share common regions and domains including a short intravesicular NH2-terminal region, a single membrane-spanning domain, a lysine- and arginine-rich region, as well as two C2 domains (C2A and C2B) located in the cytoplasmic tail [19]. Binding of Ca^{2+} to synaptotagmins via the two C2-domains transduces the Ca^{2+} signal into activation of the membrane fusion machinery, which is exerted by the interaction of the C2-domains with phospholipids and SNARE proteins [19]. Synaptotagmins 2-4 and 6-9 are expressed in pancreatic islets and β cells [20]. Synaptotagmin 7, which is co-localized with insulin granules, is thought to be the major Ca^{2+} sensor for insulin granule exocytosis [20,21].

**Metabolic amplifying pathway**

In addition to the K_ATP channel-dependent pathway that triggers insulin secretion, another pathway in GIIS augments the effect of Ca^{2+} on insulin secretion [22]. This pathway, which does not require an additional rise in [Ca^{2+}], is distinct from the hormonal and neuronal amplifying pathway, was originally referred to as the K_ATP channel-independent pathway in GIIS. However, since closure of K_ATP Channels is a prerequisite for this augmentation pathway under physiologic conditions, it is more accurately referred to as metabolic amplifying pathway. Recent studies have focused on the mitochondrial features of this pathway. As mentioned, pyruvate readily enters the mitochondrion in pancreatic β cells and is converted to oxaloacetate by PC. In gluconeogenic tissues such as liver and kidney, the activity of PC coordinates with phosphoenolpyruvate carboxykinase (PEPCK) to initiate gluconeogenesis. Although the activity of PC is also high in pancreatic β cells, β cells lack expression of both PEPCK and fructose-1,6-bisphosphatase, indicating that PC serves in roles other than gluconeogenesis. Several lines of evidence indicate that anaplerosis via PC is involved in GIIS [23,24]. Cataplerosis pathways, which provide mitochondrial metabolites to the cytosol, are also thought to be implicated in GIIS [25]. There are at least three pyruvate cycling pathways in pancreatic β cells: the pyruvate-malate shuttle, the pyruvate-citrate shuttle, and the pyruvate-isocitrate shuttle. These pathways may generate coupling factors associated with the metabolic amplifying pathway.

**Dynamics of insulin secretion**

**Biphasic insulin secretion**

Insulin release from pancreatic β cells in response to glucose is characterized by biphasic kinetics: an initial component (1st phase), which develops rapidly but lasts only a few minutes, followed by a sustained component (2nd phase) [26–28] (Figure 7.3(a)). It has been thought that the biphasic response of insulin secretion reflects primarily the dynamics of spatially and functionally distinct insulin granules. The prevailing hypothesis is that the 1st phase of insulin secretion is attributable to fusion of pre-docked granules from a readily releasable pool (RRP) that accounts for less than 5% of total granules, while the 2nd phase involves recruitment of granules from a more distant reserve pool (RP) that accounts for the great majority of total granules [27–30]. A recent study using total internal reflection fluorescence microscopy (TIRFM) found three distinct modes of insulin granule exocytosis, based on dynamics of insulin granules: a mode comprising pre-docked granules that are immediately fused to the plasma membrane by stimulation (called *old face*), another mode comprising granules that are newly recruited by stimulation and immediately fused to the plasma membrane (a docking state can hardly be detected by TIRFM) (called *restless newcomer*), and a third mode...
comprising granules that are newly recruited by stimulation, but are first paused or docked and then fused to the plasma membrane (called resting newcomer) [31]. In this model, a RRP responsible for the 1st phase is located away from the plasma membrane but is yet immediately releasable, and both 1st and 2nd phases of insulin granule exocytosis involve the resting newcomer (Figure 7.3(a)). Glucose-induced F-actin-remodeling has recently been shown to be involved in mediating the 2nd phase of insulin secretion [32] (Figure 7.3(a)).

**Exocytotic machinery**

Insulin granule exocytosis in pancreatic β cells, like synaptic vesicle exocytosis in neurons, involves several processes, including granule recruitment to the plasma membrane, docking of granules at the plasma membrane, priming of fusion machinery, and fusion of granules with the plasma membrane. However, the kinetics of exocytosis is ultrafast (a few milliseconds) in synaptic vesicles of the neuron and slow (a few hundred milliseconds) in large-dose core granules of the pancreatic β cell [33,34]. SNARE proteins critical for synaptic vesicle exocytosis are expressed in pancreatic β cells and β-cell lines [34]. This exocytotic machinery, including the t-SNAREs syntaxin 1 and SNAP-25 as well as the v-SNARE synaptobrevin/VAMP-2, function similarly in insulin granule exocytosis [33,34]. Syntaxin 1 and SNAP-25 form a cluster along the plasma membrane of pancreatic β cells [35]. The rise in [Ca\textsuperscript{2+}], triggers the formation of the SNARE complex from Syntaxin 1, SNAP-25, and VAMP-2, which promotes membrane fusion [19,34]. The SNARE complex is responsible for induction of insulin granule exocytosis in response to glucose [34].

In neurons, SNARE proteins interact with many vesicle-associated proteins including Sec1/Munc18 (SM) protein, Munc13, synaptotagmins, and complexin [34]. SM protein and Munc13 promote the assembly of SNARE proteins, and synaptotagmin and complexin control Ca\textsuperscript{2+}-dependent triggering of exocytosis [19]. SM protein associates with the closed form of syntaxins, and the closed form is presumed to prevent participation in SNARE complexes [36,37]. In pancreatic β cells Munc18-1 and Munc18c are involved in the regulation of the 1st and 2nd phases of GIIS, respectively, doing so by promoting localization of insulin granules to the plasma membrane [38]. Munc13 mediates synaptic vesicle priming by stabilizing the open conformation of Syntaxin 1, thereby allowing the formation of SNARE complexes [39,40]. In pancreatic β cells, Munc13-1 plays an essential role in the priming step in insulin granule exocytosis through its interaction with the Rab3 effector Rim2\textalpha [41]. Munc13-1 also mediates both 1st and 2nd phases of GIIS [42].

The small G-protein Rab family comprises more than 60 members [43,44]. Among them, Rab3 and Rab27a are associated with insulin granules of pancreatic β cells [45,46]. Both Rab3 and Rab27a are localized to insulin granules and function through interaction with their effector proteins Rim2\textalpha and granulophilin, respectively [41,47–49]. Rim2\textalpha plays critical roles in docking and priming steps through its interaction with Rab3 and Munc13-1, respectively [41]. The interaction of granulophilin with Syntaxin 1A/Munc18-1 is also important for docking of insulin granules to the plasma membrane [49].
Modulation of insulin secretion by various intracellular signals

GIIS is modulated by various nutrients and hormonal and neuronal inputs (Figure 7.1). Most of these hormones and neurotransmitters exert their effects on insulin secretion by binding to their specific surface receptors, which are members of the superfamily of trimeric G-protein-coupled receptors (GPCRs) [50]. Based on the properties of G-proteins, GPCRs are subdivided into different functional classes, primarily Gs, Gi/11, and Gq-protein-coupled receptors (Table 7.1). G-proteins are linked to specific signaling pathways that have multiple effects on β-cell function in the modulation of GIIS. Gs and Gi-11-proteins potentiate GIIS and may have several other beneficial effects on β-cell function [50]. Gs-protein exerts an inhibitory effect on GIIS [50]. GIIS also has been shown to be modulated by the insulin signaling pathway in rodents in vivo and islets isolated from rodents and humans [51–54].

As stimulation of insulin secretion is an essential pharmacologic strategy for treatment of type 2 diabetes, GPCR and the insulin signaling pathways in pancreatic β cells offer many attractive drug targets.

Gq-protein-coupled receptor

In pancreatic β cells, various hormones, neurotransmitters, nucleotides, and fatty acids including GLP-1 [55–57], glucose-dependent insulinotropic polypeptide (GIP) [55,57], vasoactive intestinal polypeptide (VIP) [58], pituitary adenylate cyclase-activating polypeptide (PACAP) [58], adrenalin, ATP/ADP, lysophosphatidylcholine (LPC) [59], and oleoylthanolamide (OEA) [60] activate their specific receptors (Table 7.1). These receptors when coupled with Gq-protein activate adenylate cyclase and increase cAMP production. These cAMP-increasing ligands potentiate both the 1st phase and 2nd phase of GIIS [61].

cAMP is mediated by protein kinase A (PKA)-independent as well as by PKA-dependent mechanisms, the former involving the cAMP-binding protein Epac2 (now called Epac2A) [47,62,63]. The effect of Epac2A on GIIS is mediated not only by its guanine nucleotide exchange (GEF) activity toward small GTPases but also by interactions with several other proteins, including the KATP channel regulatory subunit SUR1 [47,65,66], small G-protein Rab3 effector Rim2α [41,47,63], and Piccolo [67]. Epac2A/Rap1 signaling mediates the potentiation of the 1st phase of GIIS by cAMP [31]. It has been proposed that activation of Epac2A/Rap1 signaling increases the size of RRP and/or recruitment of insulin granules from RRP, while PKA signaling increases the size of RP and/or recruitment of insulin granules from RP (Figure 7.3(b)) [31]. Epac2A is also thought to be involved in mobilization of Ca2+ from intracellular Ca2+ stores in pancreatic β cells [68]. The effect of Epac2A is mediated by ryanodine receptors [69]. In addition, Epac2A is a direct target of antidiabetic drug sulfonylureas and is required for the effect of sulfonylureas on stimulation of insulin secretion [70–72].

PKA phosphorylation of Kir6.2, VDCC α-subunits, and GLUT2 influences their activities [73–75]. PKA phosphorylation of snapin, which interacts with SNAP25 [76], increases the interaction among insulin granule-associated proteins, thereby potentiating GIIS [77]. The subcellular localization of PKA via A-kinase anchoring proteins (AKAPs) is also critical for the stimulatory effect of cAMP-elevating agents on insulin secretion [78].

The incretin hormones GLP-1 and GIP are released from enteroendocrine L cells and K cells, respectively, in response to ingestion of nutrients [55,57]. Both hormones potentiate insulin secretion in a glucose concentration-dependent manner [79,80]. The potentiating effects of GLP-1 and GIP occur at glucose concentrations higher than 5 mM in vivo in humans [81]. Long-term treatment with GLP-1 promotes β-cell proliferation and protects from apoptosis, thereby maintaining β-cell mass in rodent pancreatic β cells and β-cell lines [55]. GLP-1 also has an anti-apoptotic effect on primary cultured human pancreatic islets [82]. However, the proliferative capacity of human pancreatic β cells and its modulation by GLP-1 is still unclear. GIP also stimulates β-cell proliferation and has an inhibitory effect on β-cell apoptosis in rodents [55]. It has been suggested

<table>
<thead>
<tr>
<th>Table 7.1 Ligands and GPCRs that modulate GIIS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ligand</strong></td>
</tr>
<tr>
<td>GLP-1</td>
</tr>
<tr>
<td>GIP</td>
</tr>
<tr>
<td>CCK</td>
</tr>
<tr>
<td>Ghrelin</td>
</tr>
<tr>
<td>Somatostatin</td>
</tr>
<tr>
<td>Glucagon</td>
</tr>
<tr>
<td>VIP</td>
</tr>
<tr>
<td>Acetylcholine</td>
</tr>
<tr>
<td>Adrenaline</td>
</tr>
<tr>
<td>Adrenaline</td>
</tr>
<tr>
<td>Noradrenaline</td>
</tr>
<tr>
<td>PACAP</td>
</tr>
<tr>
<td>NPY</td>
</tr>
<tr>
<td>Long-chain FFA (&gt;C12)</td>
</tr>
<tr>
<td>LPC, OEA</td>
</tr>
<tr>
<td>ATP, ADP</td>
</tr>
</tbody>
</table>

α2, α1-adrenoceptor; β2, β2-adrenoceptor; CCK, cholecystokinin, CCKA, CCK receptor A; FFA, free fatty acid; GCGR, glucagon receptor; GHISR, ghrelin receptor; GIP, glucose-dependent insulinotropic polypeptide; GIPR, GIP receptor; GLP-1, glucagon-like peptide-1; GLP1R, GLP-1 receptor; GPR, G-protein-coupled receptor; LPC, lysophosphatidylcholine; M3, muscarinic acetylcholine receptor subtype 3; NPY, neuropeptide Y; OEA, oleoylthanolamide; PACAP, pituitary adenylate cyclase-activating polypeptide; P2Y, P2Y-purinergic receptor; SSTR, somatostatin receptor; VIP, vasoactive intestinal polypeptide; VPAC2, VIP-pituitary adenylate cyclase-activating polypeptide receptor 2; Y1, NPY receptor 1.

Source: Adapted from Ahrén 2009 [50]. Reproduced with permission of Nature Publishing Group.
that defects in potentiation of insulin secretion by exogenous GIP is associated with reduced expression of the GIP receptor in pancreatic β cells [83–85]. Incretin-related drugs such as dipeptidyl peptidase-4 inhibitors, which block degradation of GLP-1 and GIP, and GLP-1 receptor agonists have been developed for treatment of patients with type 2 diabetes [55].

Glucagon receptors are expressed on pancreatic β cells and glucagon stimulates insulin secretion [86,87]. The binding of glucagon to its receptor activates the Gα and Gq-proteins [87]. GIIS is enhanced in pancreatic islets that are rich in glucagon compared to that from islets containing fewer pancreatic α cells [88], which indicates that glucagon is important for the insulin secretory response to glucose. However, the contribution of glucagon receptors to β-cell function remains to be established.

GPR119, which is a receptor for fatty acid, is expressed in β cells and in pancreatic polypeptide (PP) cells [89,90]. Activation of GPR119 by lysophosphatidylcholine (LPC) [59] and oleoylthanolamide (OEA) [60] increases cAMP production and stimulates insulin secretion in a glucose-dependent manner [60,89,90].

**Gq-protein-coupled receptor**

Gq-protein stimulates phospholipase Cβ to produce IP3 and DAG [50]. IP3 triggers the release of Ca2+ from the endoplasmic reticulum, whereas DAG activates protein kinase C (PKC). Activation of PKC by phorbol ester, a mimic for DAG, stimulates insulin secretion in the presence of raised intracellular Ca2+. It has been proposed that PKCe stimulates insulin secretion through the amplification of glucose metabolism [91].

Pancreatic β cells express several Gq11-protein-coupled receptors [50] including the M3 muscarinic acetylcholine (ACh) receptor (M3R) and receptors for fatty acids (GPR40), cholecystokinin (CCKA), arginine vasopressin (V1b), and extracellular nucleotides (P2Y1 and P2Y6). The M3 muscarinic receptor is involved in the regulation of insulin secretion by the vagal nerve system. ACh has a stimulatory effect on insulin secretion in pancreatic β cells through vagal nerves, an effect that mediates the cephalic phase response to food ingestion induced by the activity of efferent vagal nerves and not by the absorbed nutrients [92]. In mouse, parasympathetic and sympathetic fibers innervate islets cells including β, α, and δ cells, while in human, islets endocrine cells are shown to be barely innervated [93]. Thus, hormone secretion in human pancreatic islets may be modulated by the autonomic nervous system via sympathetic input controlling blood flow within the islets.

GPR40 is activated by fatty acids, including docosahexaenoic acid (DHA), eicosapentaenoic acid (EPA), and arachidonic acid, and potentiates GIIS [94,95]. GPR40 has been suggested to mediate the major effect of fatty acids on insulin secretion from pancreatic β cells [94,95]. GPR40 agonists also stimulate GLP-1 and GIP secretion from enteroendocrine cells, thereby amplifying GIIS from mouse pancreatic islets [96]. A GPR40 agonist has recently been developed as a potential glucose-lowering medication by stimulating insulin secretion [97].

**GIIS-activated receptors**

GIIS, noradrenaline, and ghrelin inhibit insulin secretion through the activation of PTX-sensitive Gq-protein [92,98,99], which suppresses adenylate cyclase activity. Somatostatin, which is released from pancreatic δ cells in pancreatic islets, exerts its inhibitory effect in pancreatic β cells upon binding to its specific GPCRs. There are five somatostatin receptor subtypes (SSTR1-5) [100–102]. Activation of SSTRs induces membrane repolarization and results in the reduction of action potential and Ca2+ influx [103,104]. SSTR2 is expressed predominantly in human pancreatic β cells [104].

α2-adrenoceptors activated by noradrenaline inhibit cAMP production and open K+ channels. Repolarization by K+ efflux induces the closure of VDCC and the consequent inhibition of insulin secretion [92,105].

Ghrelin has an inhibitory effect on insulin secretion in rodents and humans [99,106,107]. Circulating ghrelin is produced in X/A-like cells in rat (known as P/D1 cells in human), which is distributed predominantly in the oxyntic mucosa of stomach [108,109]. Ghrelin is also expressed in pancreatic islets in humans and released into pancreatic microcirculations. Ghrelin in β cells activates growth hormone (GH) secretagogue receptor (GHS-R) that is coupled with PTX-sensitive G-protein Gi2, decreases cAMP production, and attenuates membrane excitability via activation of voltage-dependent K+ channels (Kv2.1 subtype), consequently suppressing Ca2+ influx and insulin secretion [106,107]. It remains to be clarified how ghrelin is released from the pancreas.

**Autocrine effect of insulin on insulin secretion**

Rodent and human pancreatic β cells possess the various components of the insulin signaling system, including insulin receptor, insulin-like growth factor-1 (IGF-1) receptor, insulin receptor substrates (IRS-1 and IRS-2), phosphatidylinositol 3-kinase (PI3K), phosphoinositide-dependent kinase-1 (PDK1), and protein kinase B (PKB/Akt) [52,110–112]. It has been shown that insulin binds to the insulin receptors on the surface of β cells and induces phosphorylation of the insulin receptors and IRSs, modulating its own secretion [51,52,113,114]. This effect of insulin is involved especially in the 1st phase of GIIS in mice [52,113].

PI3K is a key component that transmits the insulin signal to downstream targets, and is shown to mediate regulated exocytosis in skeletal muscle and adipose tissues [115]. Three classes of mammalian PI3Ks have been identified based on their domain structures, differences in catalytic activity, and modes of regulation [116]. Class IA PI3K in mouse pancreatic β cells contributes to normal regulation of GIIS through the maintenance of expression levels of SNARE proteins and the control of intracellular Ca2+ levels [54,117,118]. Class II PI3K-C2α is also activated by insulin in mouse insulinoma cells and promotes GIIS via PKB/Akt1, an isomorph of PKB/Akt [119]. Knockdown of PI3K-C2α impairs insulin granule exocytosis in rat insulinoma cells [120]. The expression level of PI3K-C2α mRNA in islets of type 2 diabetic patients is decreased compared to that
in nondiabetic individuals, suggesting that downregulation of PI3K-C2α may be a feature of type 2 diabetes [120].

**Pathophysiology of pancreatic β cells**

**β-Cell mass and function in type 2 diabetes**

Type 2 diabetes is characterized by impaired insulin secretion and/or insulin resistance. However, insulin secretory capacity is the major determinant in the development of type 2 diabetes, as neither hyperglycemia nor glucose intolerance develops in insulin-resistant patients as long as sufficient insulin is secreted from β cells in a timely fashion in response to various stimuli. Impaired insulin secretion may be due to β-cell dysfunction, reduced β-cell mass, or both. Loss of β-cell mass in type 2 diabetes patients has been reported [121]. However, surgical and chemical reductions of β-cell volume induce functional adaptation of the normal β cell to prevent a rise in fasting glucose or reduction in the 1st phase of insulin secretion [122,123], suggesting that β-cell dysfunction is closely associated with the pathogenesis and pathophysiology of type 2 diabetes.

**Abnormalities in the dynamics of insulin secretion**

Abnormal dynamics of insulin secretion selective for glucose stimulation, especially the acute phase of the insulin response, is already reduced in the early stage of type 2 diabetes, patients with IGT, and their first-degree relatives [124,125]. As the early phase of insulin secretion is determined by the RRP, a reduction in its size and/or impairment of signaling for exocytosis of the insulin granules from the RRP may well develop in IGT and the early stage of type 2 diabetes. The RRP is completely depleted and the RP is markedly reduced in the fully developed stage of type 2 diabetes. Signaling mechanisms underlying exocytosis of insulin granules from the RRP and the RP may also be defective in type 2 diabetes.

Although the pulsatile nature of insulin secretion is maintained in patients with type 2 diabetes, with a number of pulses similar to that in healthy control subjects, the pulses after meals are less frequent, irregular, and have a significantly lower amplitude, resulting in a marked disruption of the dynamics of post-meal insulin secretion [126]. A loss of coordinated insulin secretory responses to oscillatory glucose infusion is found in subjects with IGT, indicating a defect in the ability of the β cells to properly sense and respond to parallel changes in the plasma glucose level [127]. Thus, abnormalities in the dynamics of insulin secretion may be an early manifestation of β-cell dysfunction preceding the development of overt type 2 diabetes.

**Conclusions**

Mechanisms of insulin secretion in pancreatic β cells have been extensively studied for decades, and many intracellular signals that trigger or amplify insulin secretion have been identified. Glucose metabolism is essential for both the triggering and amplifying pathways of insulin secretion. Besides glucose, lipid metabolism in β cells is also implicated in insulin secretion. Many molecules associated with the exocytotic machinery of insulin granules have been identified recently. Advances in technologies of imaging and mass spectrometry enable us to visualize insulin granule dynamics in living β cells and to analyze cellular proteins and metabolites in a comprehensive manner. Clarification of the molecular and cellular mechanisms of insulin secretion should provide a basis for identifying novel therapeutic targets as well as deeper understanding of the pathogenesis and pathophysiology of diabetes.

**References**

11. Brown LJ, Koza RA, Everett C, et al.: Normal thyroid thermogenesis but reduced viability and adiposity in mice lacking the


73 Leiser M, Fleischer N: cAMP-dependent phosphorylation of the cardiac-type α1 subunit of the voltage-dependent Ca2+ channel in a murine pancreatic β-cell line. Diabetes 1996;45:1412–1418.


105 Ullrich S, Wollheim CB: Expression of both α_{s} and α_{t}-adrenoceptors in an insulin-secreting cell line. Parallel studies of cytosolic free Ca^{2+} and insulin release. *Molecular Pharmacology* 1985;28:100–106.


Key points

- Insulin is crucial for survival in an environment where feeding is discontinuous and requires fine regulation of metabolism.
- Sophisticated methods have been developed over the past decades to study the multifaceted aspects of in vivo β-cell insulin secretion.
- The normal β-cell responds to intravenous glucose challenge with a complex multiphasic secretory pattern, encompassing short and transient insulin release (first phase), sustained response (second phase) and potentiation of insulin secretion for longer term glucose exposure.
- The β-cell response to intravenous glucose is potentiated when glucose is ingested orally (incretin effect); two intestinal hormones, GIP and GLP-1, mediate this phenomenon.
- Insulin secretion is further stimulated by the ingestion of proteins, as amino acids are secretagogues; fat and protein ingestion stimulate incretin hormones that potentiate insulin secretion.
- Many other factors including other gastrointestinal hormones, neuropeptides, and the nervous system are likely to contribute to controlling and modulating insulin secretion.
- Insulin resistance enhances insulin secretion, particularly in fasting conditions.
- The β-cell mass is rather variable between normal subjects; while β-cell mass is expected to impact on function, the relationship between mass and function appears to be weak.
- Peripheral insulin levels are controlled by β-cell secretion, hepatic insulin removal, and whole-body insulin clearance.
- The complex and integrated response of the β-cell is fundamental to maintain glucose levels within a relatively narrow range, despite very different environmental and metabolic conditions.

Introduction

Insulin is an ancient hormone; it emerged early in evolution since the most primitive forms of vertebrates (extant lamprey and hagfish) evolving from insulin-like peptide genes, which are expressed in all multicellular animals [1]. From an evolutionary point of view insulin facilitates survival in an environment where access to nutrients is discontinuous, erratic, and difficult, requiring the function of highly specialized tissues necessary to allow movement and appropriate reactions to external stimuli (i.e., skeletal muscle and nervous system). Insulin, in the fed state is secreted to stimulate glucose and amino acids uptake allowing the build-up of depots of glycogen, proteins and lipids, which are necessary to sustain the energy requirements during successive fasting.

It is therefore not surprising that insulin secretion is dynamic and strictly controlled by plasma glucose concentration in a feedback system that maintains rather stable plasma glucose concentrations despite erratic feeding (i.e., glucose homeostasis). The homeostatic control system is, however, asymmetric, as insulin is the only hormone counteracting hyperglycemia, while hypoglycemia is prevented by at least three other hormones (cortisol, glucagon, and adrenaline). Insulin secretion, in contrast with most of the other hormones, is also only marginally regulated by the central nervous system; the intestine through the gut hormones, reliable sensors of the feeding state, is probably the most important “distant” site of control for insulin secretion.

Insulin’s effect on the different target organs (Figure 8.1) depends on the interstitial concentrations of the hormone, which in turn depend on its secretion rate and the rate of its removal from the circulation (insulin clearance). Insulin is secreted into the portal vein and in the fasting state the portal hormone levels restrain liver supply of glucose into the systemic circulation to exactly match the need of glucose-dependent tissues (central and peripheral nervous system and red blood cells). The low levels of insulin attained in the systemic circulation in the fasting state, due also to an elevated first pass insulin clearance by the liver, allow the release of free fatty acids (FFA) from adipose tissue depots (lipolysis) so as to privilege lipids...
utilization over glucose for energetic needs. The size and the energetic density of lipid stores are in fact much greater than those of carbohydrates.

In the fed state insulin is secreted in proportion to the increments in plasma glucose levels caused by glucose entrance into the circulation from food absorption and digestion. The increased portal insulin and glucose concentrations reduce liver glucose production and promote liver glucose uptake allowing the refill of the organ glycogen depots. Systemic insulin concentration increases, because of reduced liver insulin clearance in comparison to the fasting state, resulting in an efficient inhibition of lipolysis and fatty acid release, stimulation of glucose utilization in insulin-dependent tissues (mainly adipocyte, and skeletal muscle), and replenishment of body glycogen and lipid stores. Because of these concerted effects plasma glucose increases during feeding remain limited (max 2–3 mmol L\(^{-1}\)) and return to premeal levels within 2–3 hours after the ingestion of food.

This chapter provides an account of the basic methods and main concepts concerning the physiology of in vivo insulin secretion in humans, including a brief mention of the role of insulin clearance. For each topic we will provide, when available, quantitative information and discuss its relevance to overall glucose homeostasis.

**Methods for the assessment of insulin secretion**

The functional characteristics of the β cells are evaluated in vivo using a variety of experimental approaches and measurements, which have been conceived to investigate the complex response of the β cell. These methods are described here mainly for their value in understanding the physiology of insulin secretion in response to intravenous and oral glucose; a more detailed description of the tests can be found in dedicated reviews [2,3].

**Relationship between plasma insulin concentration and insulin secretion**

The measurement of insulin concentration has been considered the most obvious approach to evaluate insulin secretion and it is still the reference method in the perfused pancreas or islet cultures. In contrast, in vivo measurement of plasma insulin concentration is believed to provide a potentially biased assessment of insulin secretion because a large fraction of the secreted insulin is removed by the liver before it reaches the systemic circulation [4]. Furthermore, insulin clearance, which is a main determinant of plasma insulin concentration, is not constant, neither in an individual nor across subjects with different characteristics (e.g. lean vs. obese, see the dedicated paragraph in this chapter). The current methodology for in vivo assessment of insulin secretion is based on the measurement of plasma C-peptide, which is co-secreted with insulin in equimolar amounts as a consequence of pro-insulin cleavage. C-peptide is virtually not extracted by the liver, and has a clearance that is considered to be more constant than that of insulin [5,6]. Insulin secretion (usually expressed in pmol min\(^{-1}\) or pmol min\(^{-1}\) per square meter of estimated body surface area) is calculated from C-peptide concentration using a mathematical operation called “deconvolution” [7]. The most common deconvolution approach
relies on standardized parameters of C-peptide kinetics, calculated from the individual anthropometric characteristics [6].

Caution should be used in the comparison across studies of insulin and C-peptide concentrations and the derived indices, as the assays are not standardized and significant differences exist between the measurement methods. An additional potential source of inhomogeneity is the conversion factor used to express the common units of insulin (μU mL⁻¹) in SI units (pmol L⁻¹); while the correct conversion factor should be 6 (1 μU mL⁻¹ = 6 pmol L⁻¹) [8], in many instances other factors have been used.

**Insulin secretion indices**

Absolute insulin secretion in itself may not be an adequate index of β-cell function, unless a standardized stimulus is provided, as insulin secretion is glucose-dependent and glucose levels vary widely under physiologic circumstances. For this reason, a common rationale of several tests is to use experimental protocols in which glucose levels are standardized, such as in the hyperglycemic clamp (described later). Under standardized conditions, appropriate indices of insulin secretion can be directly calculated. In contrast, the analysis of tests of physiologic interest such as the oral glucose tolerance test (OGTT), in which glucose levels cannot be controlled, is more problematic. Several empirical indices have been proposed largely based on the ratios between insulin concentration (or secretion) and glucose concentration [2,3]. Mathematical models have also been used, as described in the following section.

**Mathematical models**

The use of mathematical models for the interpretation of the β-cell response is as old as the initial extensive studies on β-cell function [9–11]. Modeling, indeed, allows a better quantitative evaluation of the mechanisms underlying insulin response to a challenge and allows the estimation of β-cell function also from tests in which glucose levels are not standardized.

Models have been used in the analysis of intravenous glucose tolerance test (IVGTT) [12–14] and the OGTT [15–20]. The modeling methods for the assessment of β-cell function in essence embed a mathematical description, of variable but limited complexity, of the relationship between glucose concentration and insulin secretion (or concentration). The model equations contain parameters representative of β-cell function that are estimated by fitting the mathematical model to the measured data (glucose and insulin or C-peptide concentration). There are common aspects in these models, such as the presence of a dose-response relating insulin secretion to glucose concentration, but also differences, particularly in the interpretation of the OGTT.

The models for the IVGTT are rather approximate, as in these conditions it is difficult to accurately represent the insulin response with simplified approaches. Thus, the validation of this method has been limited, in particular for the parameters quantifying late insulin response.

**β-Cell response to intravenous glucose**

Although in normal living conditions β cells are stimulated by hyperglycemia that follows glucose ingestion, the study of the response to intravenous glucose is of fundamental importance for understanding the physiology of β cells. Several tests have been developed for this purpose and this section describes the most relevant and the characteristics of insulin secretion that they reveal.

**The hyperglycemic clamp**

The most typical test for the study of β-cell response employs a brisk and sustained elevation of glucose concentration from a baseline value. In vivo, this is typically achieved with the hyperglycemic clamp technique (Figure 8.2) [21]. The β-cell response to this glucose stimulus is biphasic, with an initial insulin secretion burst lasting about 5–8 min (first-phase secretion), followed by a drop towards basal levels and then by a relatively rapid increase that persists as far as hyperglycemia is maintained (second-phase secretion).

One important physiologic feature highlighted by the hyperglycemic clamp is that the β cell responds to the increase in glucose in a proportionate manner. In other words, β-cell function, as for many other physiologic processes, can be described through a dose-response curve describing the relationship between insulin secretion and glucose concentration in steady-state conditions (Figure 8.2). This dose-response can be assessed with greater accuracy using a stepped increase in glucose concentration, the so-called graded glucose infusion test [22]. In this test, glucose is infused intravenously at increasing rates and glucose and C-peptide are measured. Insulin secretion is calculated by deconvolution of C-peptide and the β-cell dose-response is constructed by plotting insulin secretion rates at the end of each glucose infusion period versus the corresponding glucose concentration.

The β-cell dose-response is a fundamental characteristic of insulin secretion; its slope, which represents the sensitivity of the β cell to glucose, is a key β-cell function parameter. In healthy subjects insulin secretion at a basal glucose level of 5 mmol L⁻¹ is about 90 pmol min⁻¹ and increases by five-
The relationship representing the 10–15% of what is secreted per mmol L\(^{-1}\) hyperglycemic clamp. In the hyperglycemic clamp, glucose concentration is briskly elevated from the basal level using a suitable intravenous glucose infusion (a). The sharp glucose increase elicits a rapid and short-lived insulin secretion peak (first-phase secretion), followed by a drop towards basal levels and then by a relatively rapid return to a sustained value in the second half of the clamp (second-phase secretion). The hyperglycemic clamp assesses insulin secretion at two steady-state glucose levels (fasting and elevated glucose); a rudimentary \(\beta\)-cell dose-response can be determined from these data (b).

First-phase secretion becomes evident with a brisk elevation of glucose concentration. The amount of insulin secreted during the first phase is dependent on the magnitude of the glucose increase, so that even the first phase recognizes a dose-response feature [9]; in a typical +7 mmol L\(^{-1}\) hyperglycemic clamp it is around 4 nmol m\(^{-2}\) representing the 10–15% of what is secreted per hour in the second phase [23]. Though limited, the amount of insulin secreted in the first phase is relevant at least for two reasons. First, while the secretion burst is apparent with a rapid elevation in glucose concentration, the underlying secretory mechanisms appear to be active even for a more gradual rise in glucose concentration [9]. These mechanisms are likely to be responsible for a response that is anticipated compared to what would be predicted solely on the basis of the dose-response and this anticipation has relevant physiologic implications for glucose homeostasis, as discussed later. Second, first-phase insulin secretion is a very sensitive marker of early \(\beta\)-cell dysfunction [24]. Impairment of first-phase secretion is already present in subjects at risk of developing diabetes [25], and is predictive of diabetes onset [26]. Because of these relevant characteristics, assessment of first-phase secretion has been widely used.

The intravenous glucose tolerance test

A common technique for the assessment of first-phase insulin secretion is the intravenous glucose tolerance test (IVGTT, also known as FSIGTT, frequently sampled intravenous glucose tolerance test), in which glucose (typically 0.3 g kg\(^{-1}\)) is injected intravenously over a short period of time (2–3 min). The sharp rise in glucose concentration elicits a first-phase response similar to that observed in the initial part of the hyperglycemic clamp, which is commonly quantified using the so-called acute insulin response (AIR) [27]. This widely used index is typically calculated as the mean increment above baseline in insulin concentration in the first 8–10 min. An advantage of the IVGTT is its relative experimental simplicity, compared to the clamp. Moreover, similarly to the hyperglycemic clamp, parameters of insulin sensitivity can be derived as well [28]. Many of the studies concerning the relationships between insulin sensitivity and insulin secretion, discussed later, are based on this approach [29]. In contrast to AIR, late-phase insulin secretion is not commonly calculated during IVGTT testing.

When evaluated using C-peptide deconvolution, the amount of insulin secreted during the IVGTT first phase is estimated to be \(\sim 3\) nmol m\(^{-2}\), a value similar to that of the hyperglycemic clamp [30,31].

In summary, insulin secretion when assessed with intravenous glucose challenges appears to be highly dynamic with two main characteristic phases, a rapid and transient one followed by a more sustained one. These responses can be reproduced in vitro, by pancreas perfusion as well as in perfused islets [9,32]. Nonetheless, the mechanisms underlying the biphasic response remain only partially understood. It has been proposed that first-phase secretion is the consequence of the discharge of a pool of insulin granules located in the proximity of the cell membrane in response to an increase in intracellular calcium triggered by a cascade of electrochemical events generated by glucose utilization inside the \(\beta\) cell [33]. Exocytosis, however, is very sensitive to calcium levels and the typical early peaking of intracellular calcium [34] may well exert a direct contribution to this phenomenon. Second-phase secretion is also controlled by calcium, but glucose itself, independently from calcium, plays a role in sustained insulin secretion [35].

\(\beta\)-Cell response to oral glucose

The amount of insulin secreted above the basal levels during a 2-h 75 g OGTT in normal subjects is \(\sim 30\) nmol m\(^{-2}\) [31]. The insulin response to glucose ingestion is more pronounced than that elicited by an intravenous infusion achieving the same glucose levels. This difference becomes readily apparent when the
plasma glucose profile after an OGTT is reproduced by means of an intravenous infusion of glucose [36,37]. The greater response to oral glucose is referred to as the “incretin effect.”

**The incretin effect**

As shown in Figure 8.3, when glucose concentration changes during the OGTT, insulin secretion follows the glucose pattern, although insulin secretion is considerably higher after oral ingestion. This implies that the β-cell dose-response for oral glucose is shifted upwards compared to that for intravenous glucose (Figure 8.3). Estimates of the magnitude of the potentiating effect of oral glucose in normal subjects vary from study to study, partly because of the different methods used to calculate the incretin effect. Recent studies based on modeling analysis report an increase in insulin secretion of ∼1.6–1.7-fold with oral (75 g OGTT) compared to intravenous glucose administration [38,39]. The incretin effect appears to be quite variable also within subjects with normal glucose tolerance, ranging from a negligible effect to a two- to threefold amplification. The incretin effect is dose-dependent, that is, higher glucose doses elicit stronger effects; for a fivefold increase in the glucose load, the increase in incretin effect is almost twofold, from ∼1.2 to ∼2.4 [38,40].

The incretin effect is mainly attributed to the action of two hormones: glucose-dependent insulinotropic peptide (GIP) and glucagon-like peptide 1 (GLP-1) [41]. GIP and GLP-1 are secreted as glucose and other nutrients reach sections of the intestine where cells specialized in the secretion of these hormones are located (the K cells for GIP and the L cells for GLP-1) [42,43]. GIP and GLP-1 bind to specific receptors on the β cell that activate a cascade of events leading to increased insulin secretion. The action of these hormones is glucose-dependent,

---

**Figure 8.3** Incretin effect. In the test for the assessment of the incretin effect, the glucose profile observed during an OGTT (a; lighter line) is reproduced by means of a controlled intravenous glucose infusion (a; darker line). At similar glucose levels, insulin secretion is considerably increased during oral glucose ingestion (b). Consequently, the β-cell dose-response relating insulin secretion to glucose concentration is shifted upwards (c). Source: Data from [39].
that is, the increase in insulin secretion is higher when glucose concentration is higher, whereas at low glucose levels, the amplification is marginal. GIP and GLP-1 may also act indirectly on the β cell through stimulation of the nervous system, as receptors for these hormones are present in neuronal cells [44]. The secretion of GIP and GLP-1 is dependent on the amount of nutrients reaching the intestine.

Experiments employing exogenous infusion of GIP and GLP-1 have clearly shown that these hormones stimulate insulin secretion in a concentration-dependent fashion [45,46]. In particular, GLP-1 has been shown to make the β cell dose-response steeper [45], similarly to Figure 8.3. However, whether GIP and GLP-1 levels can entirely account for the incretin effect has not been fully established. The relationship between the levels of incretin hormones and the incretin effect are in fact generally weak [38,47,48]; whether this is due to limited assay precision or factors other than the hormone levels remains to be clarified.

**Insulin secretion mechanisms during an oral glucose load**

OGTT modeling analysis has enabled more careful assessment of the multiple components of insulin secretion (Figure 8.4) [19]. According to this approach, the β cell dose-response is responsible for most of the changes in insulin secretion during an OGTT. Early secretion mechanisms, which enhance insulin secretion during the initial part of the OGTT, contribute only one tenth (~3 nmol m⁻²) of total supra basal secretion (~30 nmol m⁻²) (unpublished results from [31]). Notably, the magnitude of the early insulin secretion component is similar to that calculated during the hyperglycemic clamp or an IVGTT. Therefore, based on this analysis, the contribution of early secretion phenomena is quantitatively limited during an oral glucose load as well.

The response to oral glucose is also characterized by phenomena that enhance β-cell glucose sensitivity during the test, referred to as “potentiation.” This mechanism may well account for persistent activation of insulin secretion at the end of the OGTT when glucose concentration has returned to baseline levels [20,31]. Modeling analysis predicts that in normal subjects at the end of a 2-h OGTT insulin secretion is ~70% higher compared to baseline at the same glucose levels [31].

As described earlier (Figure 8.3), a key effect of the stimulation of the incretin system is the upward shift of the β-cell dose-response, that is, an increase in glucose sensitivity. In addition, the incretin effect also enhances early secretion [38,39]. The influence of the incretin effect on potentiation is more complex. A recent study [48] has shown that the potentiation phenomena observed during an OGTT are a combination of the incretin effect and glucose-induced potentiation, a mechanism described later in this chapter.

**Insulin secretion in normal living conditions**

During a 24-h observation period in which meals are administered at the usual times, insulin secretion excursions basically follow glucose excursions during the multiple meals [49], similar to a single OGTT. However, some studies have reported that the β-cell dose-response does not remain the same during the 24 hours. In particular, in relation to the glucose levels insulin secretion is relatively potentiated during the morning meal and attenuated during the following meals and the night, to a significant extent [20,50].

**β-Cell response to nonglucose secretagogues**

**Proteins and amino acids**

The insulinotropic effect of oral proteins was first described almost 50 years ago [51–53] and recently confirmed [54–56]. After the ingestion of a small amount of proteins (30–50 g) [51,55] or a larger amount of proteins (2 g kg⁻¹) [56], plasma insulin was raised two- to threefold over baseline and remained persistently elevated for 90 or 240 minutes, respectively. In either case, blood glucose did not change, whereas both GLP-1 and GIP levels were raised threefold over fasting values [55,56].
Because the effect of the incretins is glucose-dependent [41], the insulinotropic effect of oral proteins was most likely elicited by a direct stimulation of β cells by amino acids. In support of this hypothesis, it has been shown that insulin response to different protein solutions was closely related to the increase of plasma amino acids [57–59] with large differences among the individual amino acids unexplained by the functional group [60]: phenylalanine and glycine shown to be the most potent (+30 pmol L$^{-1}$), histidine, tyrosine, and, surprisingly, arginine as neutral, and the others as intermediate (+10.5 pmol L$^{-1}$).

In vitro studies [61] have reported that the insulinotropic effect of amino acids is mostly dependent on the amino acid type, duration of exposure and concentration; furthermore, only combination of amino acids stimulates insulin secretion when added at physiologic concentrations, whereas higher concentrations of individual amino acids are required to activate insulin secretion [61]. Accordingly, the intravenous administration of a relatively small number of amino acids individually promotes insulin release while the maximum stimulus is elicited by mixed amino acids [62]. In contrast to other amino acids, homocysteine in vitro showed a dose-dependent negative effect on insulin secretion in pancreatic β cells [63].

When either amino acids or proteins were ingested with glucose, plasma insulin levels were not significant or only slightly different from those after oral glucose alone but at lower glucose concentration [52–55,59,60,64,65] suggesting a direct effect on β-cell glucose sensitivity. Interestingly, the insulinotropic effect of proteins was neither blunted by insulin resistance [66] nor by diabetes [67] and was largely explained by the enhanced GLP-1 and GIP responses.

**Lipids and free fatty acids**

The first evidence of the insulinotropic effect of nonesterified fatty acids (NEFA) dates back five decades [68–70]. Several subsequent studies, described later, led to heterogeneous results which may be explained by differences in study protocols, dose (pharmacologic/physiologic), chemical structure (level of saturation, length of the carbon chain), type of administration (acute/chronic, oral/intravenous) and in subjects’ individual features (body weight, fasting glycemia, family history of diabetes) [61,71].

In lean subjects, an acute elevation of NEFA has a minor effect upon basal insulin concentration, but significantly increased glucose-stimulated insulin secretion [72–75]; this effect is blunted in obese subjects [73,74], and it is not confirmed when NEFA are infused for 10 h [76] or ingested simultaneously with glucose [66,77]. The acute enhancement of glucose-stimulated insulin secretion seems to be dependent on the type of NEFA ingested, being more relevant after monounsaturated than after polyunsaturated and saturated NEFA [75].

Studies investigating the chronic effects of NEFA elevation yielded more conflicting results. A 24–48-h lipid infusion has been reported to reduce [78], increase [79] or not significantly change [80] glucose-induced insulin secretion. A 72-h increase in plasma NEFA concentration produced an enhanced insulin response to the hyperglycemic clamp, but no effect on basal insulin concentration and, interestingly, the effect was opposite in the subjects with family history of type 2 diabetes (T2DM) [81]. In overweight and obese subjects 24-h repeated ingestion [82] of polysaturated NEFA resulted in a greater increase of plasma insulin levels than in mono- and fully saturated, although the analysis of C-peptide profiles revealed a reduction in insulin clearance. The evidence that NEFA may affect insulin clearance has also been reported in healthy lean subjects [80,83,84] and represents another confounding factor when investigating the insulinotropic effects of NEFA.

Animal [85–87] and in vitro [70,88–90] studies show that an acute elevation of NEFA increases both basal and glucose-stimulated insulin secretion while reducing insulin clearance, whereas chronic elevation of NEFA may increase basal insulin secretion but inhibits the secretion stimulated by glucose. Different effects were elicited by different NEFA, depending on their chain length and degree of saturation [91,92].

**Slow β-cell response modes and adaptation mechanisms**

The secretion mechanisms discussed above are relatively rapid and typically suited to coping with the insulin needs of a meal. The healthy β cell can also respond with slower modes, adapting to insulin demand if required by the metabolic conditions.

**Glucose-induced potentiation of insulin secretion**

Slow response modes are unveiled by prolonged exposure to hyperglycemia or repeated glycemic stimuli. During a hyperglycemic clamp, particularly at high glucose levels (>10 mmol L$^{-1}$), the second-phase response exhibits a slow progressive rise over time [23]. When two consecutive hyperglycemic episodes are brought about, the insulin response after the second one is higher compared to the first [11]. With a prolonged (three day) infusion of glucose at a low constant rate, the insulin response assessed with a hyperglycemic clamp before and after the infusion is increased more than twofold [93]. Similarly, a prolonged glucose infusion makes the β-cell dose-response, assessed with the graded glucose infusion test, steeper [22]. When glucose is infused intravenously to mimic the response to an oral glucose test, increasing glucose doses and concentrations produces an upward shift of the β-cell dose-response, that is, insulin secretion becomes greater for the same glucose level [38]. The common denominator of these phenomena is that sustained hyperglycemia potentiates insulin secretion; this mechanism provides an additional resource to control glucose levels.
Another classical experiment showing the potentiating effect of exposure to hyperglycemia involves the use of arginine as a secretagogue. An arginine bolus elicits a burst of insulin secretion similar to that of the IVGTT. When the arginine bolus is administered in a hyperglycemic state created by a hyperglycemic clamp, the insulin secretion response is potentiated compared to the basal state [94]. In normal subjects, the magnitude of this potentiation increases almost in proportion to the glucose levels until it reaches a plateau above \( \sim 30 \text{mmol L}^{-1} \) glucose, where the response to arginine is more than fivefold the response at basal glucose. The initial slope of this curve, denoted as the glucose potentiation slope, is an index of the ability of glucose to potentiate insulin secretion [94].

Adaptation to insulin sensitivity
A widely studied adaptive mechanism of the β cell is the modulation of its response by prevalent insulin sensitivity. It is a longstanding observation that obesity is accompanied by insulin resistance and hypersecretion [95]; the widespread use of the IVGTT for the simultaneous assessment of insulin sensitivity and secretion has consolidated the notion that insulin secretion, and in particular fasting insulin secretion and the first-phase response (AIR), is inversely related to insulin sensitivity [27,29,96]. Thus, insulin-resistant subjects with normal glucose tolerance secrete more insulin to cope with the increased insulin demand.

Based on observations largely derived from the IVGTT, it has been proposed that the relationship between insulin secretion and insulin sensitivity is hyperbolic [27]. According to this paradigm, a more appropriate index of β-cell function is the product of the insulin secretion and sensitivity indices, as on the hyperbola representing normal compensation this index is constant. The product of the insulin secretion and sensitivity indices has been historically denoted as the “disposition index” [28]. The disposition index, rather than the absolute insulin secretion response, should reflect intrinsic β-cell function.

Although the experimental evidence supporting the disposition index concept has been mostly derived for AIR and the IVGTT assessment of insulin sensitivity, the paradigm has been often assumed to apply to other indices of insulin secretion. This has been criticized because it has been shown that the assumption of a hyperbolic relationship is not valid for all indices [97,98]. In the presence of inverse relationships of other kinds, the disposition index provides a biased assessment of β-cell function [98].

Physiologically more important is that β-cell glucose sensitivity obtained from an OGTT is not related to insulin sensitivity [31,99]. Therefore, while insulin resistance does upregulate fasting insulin secretion and the first-phase response, the most important β-cell function index of the OGTT is not affected. This implies that compensation for insulin resistance is of particular importance for the tuning of fasting insulin secretion and therefore fasting glucose, while the secretion mechanisms that regulate postprandial glucose are less influenced by this form of adaptation.

β-Cell response to hormones and the nervous system
As already mentioned insulin is the only hormone with a blood glucose lowering effect, while many other hormones (glucagon, cortisol, adrenaline) exert a hyperglycemic action. The changes in glucose levels elicited by these hormones obviously will be detected by the β cell, which will respond by enhancing insulin secretion. Yet, all these hormones, as well as others and the nervous system exert a coordinated direct effect on the β cell resulting in an integrated and sophisticated control network. The typical example is the intra-islet networking encompassing the entangled interaction between α, β, and δ cells [100]. Glucagon released from the α cell has a stimulatory effect on insulin secretion by the β cell, an effect that is commonly used to test residual endogenous insulin secretion in vivo [101]. The latter requires, however, pharmacologic doses of glucagon (1 mg), while the small increase in circulating insulin levels observed with glucagon infusion is more likely the result of the hyperglycemic effect of the hormone. Insulin secretion, in turn, exerts an inhibitory effect on glucagon secretion, whereas somatostatin, released from δ cells, suppresses secretion of both insulin and glucagon [102], and therefore, is believed to integrate pancreatic islet hormonal effect. More recently an interaction between somatostatin and ghrelin has been reported in regulating insulin release [103]. On top of the direct effect on the β cell, resulting in suppression of both basal and stimulated insulin secretion, somatostatin may exert an effect on insulin secretion also through modulation of a number of gastrointestinal hormones, for instance by inhibiting, with different potency, the release of gastrointestinal hormones [104].

The effects of GLP-1 and GIP on insulin secretion have already been mentioned but other gastrointestinal hormones may contribute to modulating insulin secretion. For instance, animal studies have indicated that ghrelin can counteract the insulinoinsulating effect of GLP-1 [105]. An even greater inhibitory effect has been claimed to be exerted by obestatin [106] though variable and divergent effects may be elicited by different levels of the hormone [107]. Leptin, secreted by the adipocyte, interacts with receptors located on the β cell where it causes reduced insulin gene expression and inhibits insulin secretion by a KATP channel dependent and independent pathway [108].

Contra-insular hormones such as catecholamines [109], cortisol [110], and growth hormone via the effects of IGF-1 [111] mainly exert a suppressive effect on insulin secretion, which may not be readily appreciated because of the concomitant induction of insulin resistance.
More complex appears to be the role of the nervous system. Claude Bernard first suggested the involvement of the nervous system in the regulation of insulin secretion. His hypothesis was then supported by the discovery by Langerhans that pancreatic islets were indeed highly innervated. Subsequently it was found that parasympathetic, sympathetic, and sensory nerve endings are present that can affect insulin and the β-cell function through the release of a number of neurotransmitters [112]. These include catecholamine and acetylcholine as well as neurotransmitters such as the vasoactive intestinal polypeptide (VIP), pituitary adenyl cyclase activating polypeptide (PACAP), the gastrin-releasing polypeptide (GRP), galanin, neuropeptide Y, and the calcitonin gene-related polypeptide (CGRP). The response of the β cell to nerve stimulation is the result of the balance between the stimulatory effects of the parasympathetic nerve endings and the inhibitory one of the sympathetic nerves. However, much of the information is derived from animal studies, whereas investigation into the innervation of the human islet is limited [113–115]. Acetylcholine and noradrenaline are locally released to act on cholinergic and adrenergic receptors. Acetylcholine stimulates insulin by enhancing Ca\(^{2+}\) release from intracellular depots, while activation of α\(_2\)-adrenergic receptors suppresses glucose-mediated insulin release [116] via hyperpolarization of the β-cell membrane thus contributing to rapid adaptation of insulin secretion under conditions of hypoglycemia. On the contrary a stimulation of insulin secretion is elicited by activation of β\(_2\)-adrenergic receptors. A potential role in the neuronal regulation of the activity of pancreatic islet cells appears to be exerted also by sensory fibers, including fibers containing neuropeptides GGRP and substance P, since pancreatic sensory denervation has been proposed to contribute to defective insulin secretion in Zucker diabetic animals [117].

To make the picture even more complex is the observation that a number of neurotransmitters (i.e., galanin, melatonin, melacortin, orexin, vasopressin, and so on) may also contribute to modulate β-cell function [118]. Also in this case, however, much of the information has been obtained with in vitro studies using murine islet models, while a careful profile of the human islet is still far from available. Moreover, the evaluation of the physiologic role of innervation of the pancreatic human islet poses a number of interpretation issues as stimulation or inhibition of the parasympathetic and sympathetic branches of the autonomic nervous system elicits an array of responses (metabolic changes, regional blood flow, release of multiple neurotransmitters and neuropeptides) that can ultimately result in a direct effect on the β cell. By using more selective pharmacologic agents such as tyramine, a negative effect of sympathetic activation on insulin secretion induced by i.v. arginine has been confirmed in human subjects [119].

The islet innervation may play a main role in humans by mediating the cephalic phase of insulin secretion, that is, the rapid increase in insulin secretion occurring in the first couple of minutes upon initiation of the stimulus [120]. Thus, pharmacologic denervation (parasympathetic and sympathetic inhibition by trimetaphan) results in the abolition of the cephalic phase of insulin secretion, without affecting early GLP-1 and GIP secretion [121].

More recent observations suggest that the autonomic nerves may also play a role in the synchronization of the islet ensuring simultaneous and harmonic response of the islets as a unit [122].

**Pulsatile secretion of insulin**

Insulin secretion is a very dynamic process. This characteristic is even more appreciated if the typical pulsatile secretion of the hormone is taken into consideration. Insulin oscillates with a slow ultradian periodicity (∼140 min) and a high frequency periodicity [123]. Pulse intervals have been recently calculated in a more reliable manner and shown to occur with a periodicity of 4–6 min in humans [124]. These oscillations suggest the pancreatic islet has a pacemaker function. This efficient pulsatile secretion requires formidable coordination of the secretory activity of the β cell dispersed through the 1 million pancreatic islets scattered in the 25-cm long human pancreas. This coordination process is believed to require the integrated action of intra-islet nerves, metabolites, and hormones [125]. Recent studies have shown that oscillation of intracellular calcium is synchronized with β-cell metabolism [126]. Many factors affecting insulin secretion also impact on pulsatility of insulin secretion. Both sulfonylureas and GLP-1 can enhance in vivo pulsatility [127]. Although oscillatory insulin secretion may have a role in modulating insulin action, particularly at the level of the liver, it has also been proposed recently that this pattern may have a major effect at the level of the β cell by preventing desensitization of the insulin signaling pathway that contributes to regulation of β-cell mass [128]. Pulsatility, indeed, leads to periods of low autocrine stimulation enabling the cell to set with the background concentration of circulating insulin and other growth factors. The pulse release is believed to improve release control and enhance the action of the hormone. Studies performed in normal as well as diabetic individuals have shown that less insulin is required to maintain euglycemia with pulsatile versus continuous insulin administration [129].

**β-Cell mass and function**

β-Cell mass can be accurately evaluated only through autopsy studies [130,131]. Complex in vivo tests aiming at measuring maximal secretory capacity combining different stimuli have been proposed [94,132] as an alternative. However, their ability to discriminate between defects in function and mass, as well as their feasibility, is a matter of debate [133]. As thoroughly reviewed by Robertson [134], the acute response to
an i.v. glucose bolus in normoglycemic subjects and the acute response to an i.v. arginine bolus in hyperglycemic subjects do correlate with β-cell mass. However, the severity of the defects observed with these tests in T2DM [94] contrasts with the relatively preserved mass size observed at autopsy suggesting that in vivo investigations cannot distinguish between defects of mass and function. More recently, external detection of β-cell mass through metabolic tracers has been attempted but, up to now, the reliability of these methods remains limited [135].

Autopsy studies have consistently demonstrated that the whole body β-cell mass in a normal adult man is approximately 0.9 g and displays a wide inter-individual variability (95% CI: 0.5 – 1.3 g). Assuming 1 ng per cell and a fasting insulin production of 75 pmol min⁻¹ m⁻² this means that glucose homeostasis in man relies upon “only” 900,000,000 cells, each producing approximately 90,000 molecules per minute and five- to sixfold more in the stimulated condition [33].

Obesity is a marginal factor in determining total β-cell mass as it accounts for a 0.2 g increase for any 10 units of BMI. The dimension of this cell pool expands slowly and gradually from 0.2 to 20 years of age [136] remaining relatively stable thereafter being only marginally influenced by aging [130,131]. In rodents pregnancy is associated with β-cell proliferation until mid gestation to drop to the nonpregnant stage after parturition [137].

The reasons for the large individual variability in β-cell mass in adults ranging from 500 to 1500 million cells is not fully understood but a genetic influence cannot be ruled out as suggested by the observation that several type 2 diabetes susceptibility genes are involved in β-cell development (PDX1, PTF1A, or HNF1B) and apoptosis (INS, HNF4A, EIF2AK3, WFS1, or FOXP3) [138].

The workload imposed on each single cell will vary in inverse proportion to the initial asset and, as a corollary, the β-cell functional reserve must be wide enough to guarantee normal glucose homeostasis in most individuals. As evident from autopsy [130,131] and pancreatic surgery [139] studies, this functional reserve is approximately twofold, with hyperglycemia usually ensuing for reduction of >50% of the original β-cell mass. After hemi-pancreatectomy similar glucose values are achieved in response to either oral or i.v. glucose in spite of 50% lower insulin response suggesting that the reduction of β-cell mass translates into — or mimics — a diffuse and severe β-cell dysfunction. The small impact on glucose tolerance is likely to depend on the simultaneous reduction of glucagon [140] and/or increase in insulin sensitivity. Interestingly, in a similar experimental setting the C-peptide/glucose ratio was found to correlate with the residual β-cell mass [141]. However, the explained variance was less than 50% and it was present only when the subjects with diabetes, those with impaired and those with normal glucose tolerance were evaluated altogether. These observations suggest that a reduction in β-cell mass can translate into impaired function. However, such an effect tends to become evident only in extreme conditions (i.e., existing insulin resistance), whereas, in a more physiologic setting, the relationship between mass and function is not present or is rather loose.

The evidence of heterogeneous, and sometimes selective, β-cell dysfunctions in subjects with mild impairments in glucose homeostasis [25], points to a minor impact of β-cell mass on physiologic glucose homeostasis. This is also supported by the demonstration that mild-to-severe degrees of β-cell dysfunctions can be reversed by intensive blood glucose-lowering therapy [142] or bariatric surgery [143]. In an elegant study performed in non-diabetic subjects undergoing partial pancreatectomy [144] the same ~50% reduction of the β-cell mass, resulted in different degrees of β-cell dysfunction depending on the subject’s insulin sensitivity, underscoring the discrepancy between mass and function. Finally, in a very recent study in a rodent model of T2DM [145] a significant fraction (75%) of β cells were found to be present but not producing insulin in response to glucose, confirming the dissociation between function and mass.

### Insulin clearance

Insulin clearance is usually expressed in two ways, depending on the site of entry of the hormone into the circulation: (i) exogenous (or peripheral) insulin clearance, which can be determined experimentally through a euglycemic glucose clamp as the ratio between exogenous insulin infusion and arterial insulin concentration at steady-state, and (ii) endogenous (or prehepatic) insulin clearance, which is calculated as the ratio between endogenous insulin secretion and arterial insulin concentration at steady-state. The latter is an important determinant of plasma insulin concentration in physiologic conditions; however, its direct experimental determination is hampered by the difficulty in gaining access to the portal vein where endogenous insulin is secreted. Thus, its determination typically rests on the indirect calculation of insulin secretion through C-peptide deconvolution. Assembling data from different studies and our own data, in Figure 8.5 we provide estimates of the values of whole body and major organs insulin clearance in a subject both in the fasting state and during a 2-h OGTT; the two conditions will be discussed separately.

### Insulin clearance in fasting conditions

In quantitative terms the fraction of portal insulin that is removed by the liver in its first pass is approximately 65%, ranging between 50 and 70% [50,146–148]. Once into the systemic circulation, insulin is cleared again by the liver with a similar efficiency and to a lesser extent by the skeletal muscles and the kidneys. The overall contribution of the liver (first pass plus recirculation) is therefore dominant (approximately 90%).

Information on the biologic variability of endogenous fasting insulin clearance is lacking and can be indirectly inferred from the data on exogenous insulin clearance. Exogenous insulin
clearance displays a large inter-individual variability, which is chiefly (~50%) explained by genetic factors [149,150] and, to a lesser extent (~20%), by the negative effect of abdominal obesity and liver fat [151], and insulin resistance [152]. The role of insulin clearance in the maintenance of glucose homeostasis is commonly interpreted as compensatory, with lower values being observed in those who are more insulin-resistant and would take advantage of increased insulin levels. This may not hold true since experimental animals in which insulin clearance was reduced by knocking out the insulin degrading enzyme [153], after a transient mild improvement in glucose metabolism developed insulin resistance and diabetes. Similarly, fasting hyperinsulinemia and mild degrees of insulin resistance developed in pancreas-transplanted patients with organ venous drainage into the systemic circulation compared to those with portal drainage [154]. In physiologic conditions there is no convincing evidence that endogenous insulin clearance plays a role in the control of fasting glucose homeostasis [155].

**Insulin clearance in the fed state**

In the fed state endogenous insulin clearance is lower compared to fasting (Figure 8.5) as a consequence of the saturation of hepatic insulin removal for elevated insulin concentrations and possibly of other unknown factors. The contribution of the muscle is reduced [156] while that of the kidney is increased [147] although together they still contribute no more than 20% to overall insulin clearance. The 25–30% reduction with respect to the fasting rate displays a wide inter-individual variability (from
10 to 40%), which is largely, though not entirely, explained by the different insulin concentrations (lower clearance for higher levels). A recent study [157] has revealed the mechanism for the saturation and a source of variability: mice lacking the transporter (ZnT8), which enriches the insulin vesicles in zinc, demonstrate that this ion is responsible for the inhibition of the clathrin-dependent insulin endocytosis in the liver and subjects with a single nucleotide polymorphism in the gene encoding for this transporter also display an increased insulin clearance during an OGTT. Little information is available on the nongenetic factors that influence the reduction in insulin clearance observed in the fed state. A role of the central nervous system is possible [158] as well as signals other than GLP-1 from the gut [159] to the liver via nitric oxide generation [23,160].

The presence of reduced insulin clearance in all the conditions of hyperinsulinemia/insulin resistance (diabetes, obesity, fatty liver) has led to the conclusion that the changes in clearance are compensatory: they guarantee higher peripheral levels at similar secretion rates. This association, however, is largely dependent on the saturation phenomenon and it is thus very difficult to extrapolate the contribution of insulin clearance to glucose homeostasis independently of the prevailing plasma insulin levels. On the other hand, there is evidence that insulin clearance can influence glucose tolerance independently of other factors. For instance, an acute 30% increase in insulin clearance, produced by systemic nitric oxide inhibition, is able to deteriorate glucose tolerance in normal subjects [23]. Furthermore, surgical alteration of insulin clearance in pancreas-transplanted patients with portal or systemic pancreatic drainage [154] produces homeostatic changes that are independent of β-cell function. Therefore, in the fed state insulin clearance, by modulating insulin levels, appears to contribute to glucose homeostasis in concert with, and not in response to, insulin sensitivity and secretion.

**β-Cell function and glucose homeostasis**

Insulin secretion, together with the sensitivity to insulin of glucose-utilizing tissues, is a key player in glucose homeostasis. What is relevant for glucose homeostasis is not the absolute insulin secretion levels but insulin secretion relative to glucose, as typically assessed by the β-cell dose-response during the physiologic condition of oral glucose ingestion (Figure 8.3). Thus, β-cell glucose sensitivity is strongly inversely associated to mean glucose levels during a standard OGTT and explains, together with insulin sensitivity, a substantial proportion of the variability in glucose levels [31,99].

It should be recalled that two distinct mechanisms determine glucose sensitivity: the ability of the β cell to respond to intravenous glucose and the incretin effect. Both these mechanisms contribute to glucose homeostasis in a significant manner [38,161].

The different modes of response of the β cell exert a specific role in the regulation of glucose levels. One important aspect is the control of fasting glucose levels, which is achieved by a modulation of fasting insulin secretion. In healthy subjects, insulin resistance does not produce relative fasting hyperglycemia, as the fasting secretory tone is properly upregulated, to the extent that fasting glucose is not related to insulin sensitivity [31]. Notably, upregulation of fasting secretion occurs through an upward shift of the β-cell dose-response, induced by insulin resistance. This mechanism is specific for the basal condition, as it does not affect the dose-response slope, that is, glucose sensitivity [31].

The role of the early secretion mechanisms is more difficult to ascertain. The ability of the normal β cell to anticipate the secretory response is reputed to provide a “priming” mechanism that compensates for the relatively slow action of insulin in muscle, or possibly restrains glucose production more efficiently [162]. Indeed, the estimate of early secretion obtained by modeling analysis has been shown to be related to glucose levels [31]. A more direct proof of the relevance of an anticipated secretory response comes from a study evaluating the effects on glucose levels of two insulin profiles with the same mean value but different timing: the anticipated profile produces lower glucose than the delayed one [163].

Potentiation phenomena also have a role in glucose tolerance. Enhancement of insulin secretion at the end of an OGTT compared to the beginning has been shown to be related to 2-h glucose [31,164]. It is also likely that the potentiation phenomena observed during the 24 hours contribute to a more efficient control of glucose tolerance.

**Conclusions**

The β cell is an extraordinary organ. Only one gram of tissue exerts a function that is fundamental for living. The β-cell response exhibits a high degree of sophistication and can finely tune insulin secretion to cope with the metabolic needs. Appropriate control of glucose levels, indispensable for healthy living, is almost entirely ensured by this small but powerful organ. β-Cell dysfunction may therefore easily result in alterations to the multiple biologic action of the hormone. As discussed in depth in Chapter 24, β-cell dysfunction is a key feature of diabetes. Comprehension of the mechanisms controlling such a fine and articulated function is not straightforward but gaining an understanding of the as yet unresolved questions in order to pave the way for new and more effective cures for this widespread disease is of prime importance.

**References**


145 Do OH, Low JT, Gaisano HY, Thorn P: The secretory deficit in islets from db/db mice is mainly due to a loss of responding beta cells. Diabetologia 2014;57:1400–1409.


CHAPTER 9
Neuropeptides and islet hormone secretion

Bo Ahrén
Department of Clinical Sciences Lund, Lund University, Lund, Sweden

Key points

• Autonomic nerves, being parasympathetic, sympathetic or sensory, innervate the pancreatic islets.
• When these nerves are activated they release neurotransmitters, which diffuse a short distance to the islet β and α cells and activate specific receptors resulting in stimulation or inhibition of islet hormone secretion.
• Islet nerves also innervate islet blood vessels, and may therefore also affect islet function through changes in blood flow.
• Besides the main neurotransmitters, acetylcholine in parasympathetic nerves and noradrenaline in sympathetic nerves, the islet autonomic nerves also release neuropeptides.
• The main islet nerve neuropeptides are pituitary adenylate cyclase-activating polypeptide (PACAP), vasoactive intestinal polypeptide (VIP) and gastrin-releasing polypeptide (GRP) in parasympathetic nerves, galanin and neuropeptide Y (NPY) in sympathetic nerves and calcitonin gene-related peptide (CGRP) and substance P in sensory nerves.
• The parasympathetic nerves, through their neurotransmitters stimulate insulin and glucagon secretion, whereas the sympathetic nerves, through their neurotransmitters inhibit insulin secretion and stimulate glucagon secretion.
• The physiologic function of the autonomic nerves is still not established, although they may be of importance for the islet hormone response to the early phase of feeding and stimulation of glucagon secretion during hypoglycemia.
• Several key questions need to be solved in future research for a thorough understanding of islet neurobiology, the impact of the islet autonomic nerves for physiology and pathophysiology, and the relative contribution to such impact by the neuropeptides.

Introduction

The traditional view is that islet hormone secretion is mainly regulated by circulating nutrients (glucose, amino acids, free fatty acids) as well as the gut incretin hormones (glucagon-like peptide 1 (GLP-1) and glucose-dependent insulinotropic polypeptide (GIP)). However, in certain circumstances regulation of insulin and glucagon secretion is also dependent on the autonomic nerves which innervate the islets [1]. These nerves belong to the parasympathetic, sympathetic, and sensory branches of the autonomic nervous system (Figure 9.1(a)). Their nerve terminals traverse in close apposition to the islet endocrine cells and contain in their vesicles both the classical neurotransmitters (acetylcholine and noradrenaline) and several neuropeptides, which are released from activated nerves (Figure 9.1(b)). The neurotransmitters then diffuse a short distance to the islet β- and α cells and activate specific receptors resulting in stimulation or inhibition of islet hormone secretion (Figure 9.1(c) illustrates the parasympathetic-islet complex). The islet nerves also innervate the islet blood vessels, and may therefore also affect islet function through changes in blood flow.

Islet parasympathetic nerves

The parasympathetic nerves innervating the islets originate in the pancreatic ganglia, whose activity is controlled by preganglionic nerves, originating in the dorsal motor nucleus of the brain [1]. In addition, the adjacent duodenum projects parasympathetic fibers to the pancreatic ganglia [2]. The preganglionic nerve fibers form classical nerve terminals and synapses within the ganglia where they release acetylcholine to activate nicotinic receptors on the postganglionic neurons. The nerves have small vesicles which contain acetylcholine and large dense core vesicles which contain neuropeptides. Both acetylcholine and neuropeptides are released from small varicosities in the nerves, from which they diffuse to the islet endocrine cells and activate specific receptors which cause stimulation of both insulin and glucagon secretion. A main physiologic importance of this regulation is the early phase
of feeding, both during the cephalic phase of meal-induced insulin secretion, which is the release in anticipation of the meal, and during the first few minutes after food is ingested but before significant amounts of nutrients have reached the islets to stimulate insulin secretion directly [3]. Another physiologic importance of the parasympathetic islet nerves may be, together with the sympathetic nerves, the stimulation of glucagon secretion during hypoglycemia [4].

The classical mechanism invoked to explain stimulation of insulin and glucagon secretion during vagal nerve activation is the release of the neurotransmitter acetylcholine. The released acetylcholine diffuses to the endocrine cells and activates muscarinic receptors (mainly the M1 and M3 subtypes of the muscarinic receptor) which results in stimulation of insulin and glucagon secretion [5]. The intracellular second messenger system induced by muscarinic receptor activation involves phospholipase C (PLC) with increased formation of inositol-1,4,5-trisphosphate (IP$_3$) and liberation of calcium from intracellular stores; this occurs in conjunction with formation of diacyl glycerol (DAG) and subsequent activation of protein kinase C (PKC) [6,7]. Another mechanism is phosphorylation-/arrestin-dependent coupling of the muscarinic M3 receptor to protein kinase D1 [8].

The critical importance of the cholinergic mediation of islet hormone secretion for glucose homeostasis is evident from studies using mice with either β-cell deletion of the M3 receptors or β-cell overexpression of M3 receptors [8]. The β-cell M3 knockout mice have defective insulin secretion and glucose intolerance, whereas β-cell M3 receptor overexpressing mice have increased insulin secretion and enhanced glucose tolerance. The mice with β-cell overexpression of M3 receptors are also resistant to diet-induced glucose intolerance and hyperglycemia [8].

Neuropeptides contained in islet parasympathetic nerve terminals may also contribute to vagally induced stimulation of insulin and glucagon secretion, since in the dog insulin secretion after vagal nerve activation is not abolished by atropine, which blocks the muscarinic receptors. However, it is abolished by hexamethonium, which blocks the nicotinic receptors needed to activate postganglionic nerves [9]. The neuropeptides localized to islet parasympathetic nerve terminals and therefore potential mediators of the islet noncholinergic effects are gastrin-releasing polypeptide, vasoactive intestinal polypeptide, and pituitary adenylate cyclase-activating polypeptide [9].

**Gastrin-releasing polypeptide (GRP)**

GRP is the mammalian homologue of the amphibian peptide bombesin. It consists of a 27-amino acid residue, which is α-amidated in its C-terminal methionine. The peptide is highly conserved during evolution, having an identical N-terminal end to bombesin, and the human and porcine forms of the peptide differ in only two residues. GRP is widely distributed
in mammalian tissues, with particular abundance in the lung, the central nervous system, and the gut. In the pancreas, GRP is localized to islet nerves [1] and nerve terminals in the pancreatic ganglia, and is released from the isolated pig pancreas during vagal nerve activation [10]. GRP stimulates insulin secretion as shown in experimental studies in mice, pigs, and dogs [10–12]. This stimulation is mediated by activation of both β-cell and ganglionic GRP receptors, as evident from findings in vivo that the ganglionic antagonist hexamethonium and the muscarinic antagonist atropine both partially inhibit GRP-stimulated insulin secretion [1,13]. A main mechanism underlying the direct β-cell stimulation of GRP is increased cytoplasmic calcium due to stimulation of both calcium uptake from extracellular space and release from intracellular stores [14]. However, PKC [15] and phospholipase D (PLD) may also contribute to GRP-induced insulin secretion [1]. Hence, activation of several different signaling mechanisms may together contribute to the insulinotropic action of GRP.

Bombesin-like peptides activate several subspecies of receptors, but it is the GRP receptor (GRPR) that is the receptor subtype responsible for GRP-stimulated insulin secretion [16]. To study the physiology of GRP and islet function, mice with genetic deletion of the GRPR have been developed. These mice exhibit impaired insulin secretion in response to autonomic nerve activation by neuroglycopenia induced by 2-deoxyglucose (2-DG) [16]. 2-DG is known to activate the autonomic nerves, thereby increasing insulin secretion in mice by an effect that is counteracted by ganglionic blockade by hexamethonium [1]. Insulin secretion in response to oral glucose is also impaired in GRPR-deleted mice, indicating involvement of GRP in neural stimulation of insulin secretion after meal ingestion [17].

In contrast to the well-documented stimulation of insulin secretion by GRP, its potential effect on glucagon secretion is not established. One study demonstrated stimulation of glucagon secretion by GRP in mice [11], but other studies in dogs showed no effect on glucagon secretion [12]. Overall, GRP’s role seems more related to the insulinotropic than the glucagonotropic action of the parasympathetic nerves.

**Vasoactive intestinal polypeptide (VIP)**

VIP was originally isolated from the porcine small intestine in the early 1970s and found to consist of 28 amino acid residues, with a C-terminal α-amidation showing structural similarities to other members of the glucagon superfamily of peptides [18]. VIP is formed by processing of 149-amino acid proVIP and is highly conserved as illustrated by human, porcine, and rat VIP being identical and differing from guinea pig VIP by only two amino acids.

VIP nerves are localized both to the central nervous system and to the respiratory, gastrointestinal, and genitourinary systems. This neuropeptide exerts a variety of actions, including relaxation of smooth muscle cells and stimulation of exocrine and endocrine secretions. In the pancreas, a neural VIP network surrounding the islets has been demonstrated both in humans and experimental animals and it has also been demonstrated that islet VIP nerves show particular association with the centrally located insulin-producing cells [19,20]. VIP has also been demonstrated in ganglionic nerve cell bodies, the source of the parasympathetic nerves that innervate the islets. Like GRP, VIP is also released from the pancreas when the vagal nerves are activated [20]. VIP is a powerful stimulator of insulin secretion as demonstrated after exogenous administration in vivo and in vitro [21]. Similarly, transgenic mice overexpressing the VIP gene in the islet β cells show increased insulin secretion and increased glucose tolerance with a reduction of circulating glucose [22].

Two VIP binding receptors exist and because they also show affinity for pituitary adenylate cyclase-activating polypeptide (PACAP) they are called the VPAC1 and VPAC2 receptors (VPAC1R and VPAC2R) [23]. They are both G-protein-coupled receptors having seven transmembrane domains. In situ hybridization using probes for the messages of VPAC1R and VPAC2R has demonstrated that VPAC2R but not VPAC1R is expressed in rat and mouse islets as well as in insulin-producing clonal cells [24]. This suggests that it is VPAC2R that transmits the VIP signal to islet endocrine cells. Through activation of VPAC2R, VIP activates adenylate cyclase, which increases the formation of cyclic AMP (cAMP), as has been demonstrated in islet cells [25]. cAMP in turn activates PKA, which enhances exocytosis in both direct and indirect manners. VIP also stimulates glucagon secretion as demonstrated in several species [21]; however, the mechanism underlying the glucagonotropic action of VIP has not been established.

**Pituitary adenylate cyclase-activating polypeptide (PACAP)**

PACAP exists in two forms, 27 and 38 amino acids, both of which, like VIP, belong to the glucagon superfamily. VIP and PACAP show 68% identity. The C-terminal 11 amino acids of PACAP38 are cleaved to produce PACAP27, with the main form being PACAP38 [26]. PACAP is found mainly in nerves of the central nervous system, the lungs, and the gastrointestinal tract. It is also found in nerve terminals and their neuronal cell bodies in pancreatic ganglia, as demonstrated both in mice and rats [26]. The localization of PACAP to autonomic nerve terminals in humans has not yet been established. The islet PACAP nerves are thought to be mainly parasympathetic nerves, because PACAP is co-stored with VIP in some islet nerves, is released from the pancreas during electrical activation of the vagal nerves, and pancreatic PACAP content is reduced by vagotomy.

PACAP potently stimulates insulin secretion as demonstrated both in vivo and in vitro [21,26]. It also augments glucose-stimulated insulin secretion in healthy volunteers.
upon exogenous administration, showing that PACAP is an insulinotropic neuropeptide in humans [27]. PACAP stimulates insulin secretion in a glucose-dependent manner via activation of adenylate cyclase, increased formation of cAMP, and activation of PKA [26]. PACAP also increases the cytoplasmic concentration of calcium and sodium through opening of membranous calcium and sodium channels, and stimulates a distal effect on the exocytosis machinery. The potency of PACAP to stimulate insulin secretion, which exceeds that of VIP, is likely due to the diversity of intracellular signals it activates to stimulate islet β cells.

There is a close relation between PACAP and VIP, which is evident by the structural similarity of the peptides and by the affinity of both peptides for VPAC1R and VPAC2R. However, PACAP also stimulates a third subtype of receptors, which is specific for PACAP. This receptor, the PAC1 receptor (PAC1R), like the two VPAC receptors, is G-protein-coupled and has seven transmembrane domains [23]. Gene expression and in situ hybridization studies have shown that PAC1 receptors and VPAC2 receptors are expressed in islets [26]. To study the physiology of PACAP, PAC1R knockout mice have been made [28]. These mice are viable, develop normally and have normal baseline glucose and insulin levels as well as normal pancreatic insulin content. The insulin response to PACAP is reduced by 50% in PAC1R knockout mice, showing that a main mediator of the PACAP-induced insulin secretion is the PAC1 receptor. A most interesting phenotype in the PAC1R-deleted mice is that they exhibit an impaired insulin response to glucose, both in vivo and in vitro. This suggests that PAC1R contributes both to PACAP- and glucose-induced insulin secretion, that is, PACAP is of physiologic importance for glucose-stimulated insulin secretion. In regard to these PAC1R-deleted mice, it is important to remember that expression of VPAC1R and VPAC2R is still intact, which likely allows PACAP to still exert some stimulation of insulin release and may explain why PACAP-induced insulin secretion is not abolished, but reduced only by approximately 50% in the PAC1R-deficient mice. Therefore, studies on VPAC1R- and VPAC2R-deleted mice are important for a more complete understanding of the insulinotropic role of PACAP. PACAP knockout mice are unlikely to be useful in this regard since they exhibit high mortality and have a marked psychomotor phenotype although their islet function has not been studied [26].

Because of its localization to parasympathetic nerves in the pancreas and release from the pancreas after vagal nerve activation as well as its potent stimulation of insulin secretion, PACAP may be of physiologic relevance for neural regulation of islet function before and during meals. Results showing that the insulin response to oral glucose in mice is inhibited by a PACAP receptor antagonist and impaired in PAC1R-deleted mice support this hypothesis [26,28]. However, further studies are needed to determine its exact role.

PACAP has also been shown to stimulate glucagon secretion in several species, including humans [21,26,27]. It therefore mimics both the stimulation of insulin and glucagon secretion induced by parasympathetic nerve activation. The glucagon response to insulin-induced hypoglycemia has also been shown to be impaired in mice with genetic deletion of PAC1 receptors [29], which may suggest that PACAP contributes to neurally induced glucagon counterregulation.

### Parasympathetic effects and mediation

As reviewed earlier, activation of the parasympathetic nerves stimulates both insulin and glucagon secretion, and this may be of particular relevance for the cephalic and early meal phases of insulin secretion as well as for the glucagon response to hypoglycemia. These effects may be mediated by the classical neurotransmitter acetylcholine and by the neuropeptides confined to these nerves (GRP, VIP, and PACAP). These four neurotransmitters are all potent stimulators of insulin and also, with the exception of GRP, glucagon secretion, and therefore mimic effects of parasympathetic activation on islet hormone secretion. In classical neurobiology, several criteria need to be met before a candidate neurotransmitter, such as these neuropeptides, can be proven to be a physiologic neurotransmitter [30] (Table 9.1). These criteria include that the potential neurotransmitter is localized to the nerves, is released by activation of the nerves, its effects mimic that of the nerves and that inhibition of its effect alters the responses to nerve activation. Some of these criteria have been met for GRP, VIP, and PACAP, and therefore some of them are strong candidates for physiologic neurotransmitters (see Table 9.2). Currently, convincing evidence has been presented for VIP and PACAP and evidence also exists for GRP. However, further studies are still required to firmly establish the involvement of GRP, VIP, and PACAP in parasympathetic control of islet function.

### Islet sympathetic nerves

Islets are densely innervated by sympathetic nerves, as demonstrated by fluorescence microscopy, electron microscopy, and immunocytochemistry [1]. The islet sympathetic nerves are

**Table 9.1** Four criteria for a proposed function of neurotransmitter in the islets [30]

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Criterion 1</strong></td>
<td>The potential neurotransmitter mimics the effects of nerve activation on insulin or glucagon secretion when exogenously added to islets (direct stimulation)</td>
</tr>
<tr>
<td><strong>Criterion 2</strong></td>
<td>The neurotransmitter is localized to islet nerve cells</td>
</tr>
<tr>
<td><strong>Criterion 3</strong></td>
<td>The neurotransmitter is released from the pancreas when the nerves are activated</td>
</tr>
<tr>
<td><strong>Criterion 4</strong></td>
<td>Interference with effect of the potential neurotransmitter (receptor antagonist or genetic deletion of neurotransmitter or receptor) alters response to nerve activation</td>
</tr>
</tbody>
</table>
postganglionic nerve fibers with most nerve cell bodies located in the celiac ganglion and some in the paravertebral sympathetic ganglia (Figure 9.1(a)). Electrical activation of the sympathetic nerves inhibits insulin secretion and stimulates glucagon secretion [31]. The effects of sympathetic nerve activation on insulin and glucagon secretion are partially mediated by the classical neurotransmitter noradrenaline through activation of α-adrenoreceptors, which in turn results in inhibition of adenyl cyclase, with reduced formation of cAMP. However, hyperpolarization with closure of the voltage-sensitive calcium channels and a direct exocytotic action may also be involved.

In addition to noradrenaline, neuropeptides also contribute to sympathetically induced inhibition of insulin secretion. The evidence for this is twofold. First, combined α2- and β-adrenoreceptor blockade by yohimbine and propranolol does not prevent the effect of sympathetic activation to inhibit basal insulin secretion. Second, exogenous noradrenaline does not mimic the effect of sympathetic activation to inhibit basal insulin secretion [31]. Two sympathetic neuropeptides, galanin and neuropeptide Y, may mediate these nonadrenergic effects because they (a) are localized to both the islet sympathetic nerve terminals and their neuronal cell bodies in the celiac ganglion, (b) are released from the pancreas in response to electrical activation of the sympathetic nerves, and (c) each inhibit insulin secretion as demonstrated both in vivo and in vitro [1].

Galanin

Galanin was isolated from the porcine gastrointestinal tract in 1983 and named because of its glycine C-terminus and alanine N-terminus [32]. In most species it is 29 amino acids in length and amidated in its C-terminus. However, the human form is 30 amino acids long and is not C-terminally amidated.

Galanin is a widely distributed neuropeptide both in the central and peripheral nervous systems. In islets, dense galanin innervation was first demonstrated in dogs, and later in other species, including humans [30,32]. However, species differences exist; for example, islet galanin innervation is more abundant in dogs and cats than in common laboratory animals like rats and mice [32]. In dogs, islet nerves that contain galanin also contain tyrosine hydroxylase as do their neuronal cell bodies in the celiac ganglion [30]. Thus, islet galanin nerves appear to be sympathetic. This hypothesis is supported by the substantial, although not total, loss of pancreatic galanin nerves that develops after chemical sympathectomy with 6-hydroxydopamine. Further evidence comes from the dog, whose pancreatic galanin is released after sympathetic nerve activation, and from the mouse, whose pancreatic galanin content is reduced by exercise stress [30]. Therefore, although galanin may occasionally also be found in certain other nerves, most data suggest that islet galanin is a sympathetic neuropeptide.

Galanin has been demonstrated to potently inhibit basal and stimulated insulin secretion both in vivo, as demonstrated in dogs, rats, and mice, and in vitro, as shown in isolated rodent and human islets and in perfused rat pancreas. Galanin also stimulates glucagon secretion [32]. The actions of galanin have been shown to be transmitted by signals induced by activation of three different galanin receptor subtypes, called GalR1, GalR2, and GalR3, respectively. All three are G-protein-coupled receptors encoded by different genes [32]. The GalR1 form is expressed in insulin-producing cells and therefore most likely the mediator of the inhibitory effect of galanin on insulin secretion [33].

The mechanism underlying this inhibitory action involves opening of ATP-sensitive potassium channels and inhibition of the cellular uptake of calcium by closure of voltage-sensitive calcium channels through hyperpolarization, although also inhibition of the formation of cAMP through involvement of a pertussis toxin-sensitive G-protein, and through actions on proteins which are of critical importance for the exocytosis of secretory granules have been suggested [32–35].

Functional studies have demonstrated that galanin is released from the dog pancreas during sympathetic nerve activation in a quantity that is sufficient to mimic the inhibition of insulin secretion induced by sympathetic nerve stimulation [30]. These data therefore suggest that galanin

---

**Table 9.2** Application of the criteria from Table 9.1 for neurotransmitters in the parasympathetic islet nerves to mediate the stimulation of insulin and glucagon secretion which is achieved by activation of the parasympathetic nerves

<table>
<thead>
<tr>
<th>Neuropeptide</th>
<th>Criterion 1</th>
<th>Criterion 2</th>
<th>Criterion 3</th>
<th>Criterion 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRP</td>
<td>Yes</td>
<td>No∗</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>VIP</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>PACAP</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

∗Not consistent in literature.
helps mediate sympathetically induced inhibition of insulin secretion. A similar conclusion comes from a study in mice in which two minutes of swimming causes a 50% inhibition of glucose-stimulated insulin secretion [36]. When galanin was immunoneutralized in these mice by pretreatment with a high-titer galanin antiserum, swimming no longer inhibited glucose-stimulated insulin secretion, suggesting that galanin mediated this exercise-induced inhibition of insulin secretion. Further support comes from galanin knockout mice where the inhibitory effect of sympathetic nerve activation of insulin secretion is impaired [37]. Although a galanin receptor antagonist, galantide, exists [38], studies using this approach to prove a role for galanin in sympathetically mediated inhibition of insulin secretion remain to be performed.

**Neuropeptide Y (NPY)**

NPY was isolated from the porcine brain in 1982 and shown to be of a 36-amino acid residue with tyrosine at both N- and C-termini. It has structural homology with peptide YY (PYY) and pancreatic polypeptide (PP) and is widely distributed in the central and peripheral tissues. In most tissues the peptide is localized to perivascular adrenergic nerve fibers. In the pancreas, NPY-containing nerve terminals have been demonstrated around vessels, in the exocrine portion of the gland, as well as in islets in dogs, pigs, and mice [1,39,40]. NPY is co-localized with tyrosine hydroxylase in most islet nerves, suggesting that it is mainly a sympathetic neuropeptide. This hypothesis is supported by the localization of NPY to nerve cell bodies of the celiac ganglion, which also stain for tyrosine hydroxylase [1,40]. However, some islet nerves that do not contain tyrosine hydroxylase also express NPY, suggesting an additional nonsympathetic function for islet NPY [40]. NPY is released from the pancreas during nerve activation and most studies suggests that NPY inhibits glucose-stimulated insulin secretion, that is, it exerts sympathetic-like effect, both *in vivo* and *in vitro* [1,39–41].

Several NPY receptors exist, all of which are G-protein-coupled with seven transmembrane domains. In insulin-producing cells, it is the Y1 receptor subtype (Y1R) that mediates islet actions of NPY [41]. The mechanism underlying NPY’s inhibition of glucose-stimulated insulin secretion involves inhibition of adenylate cyclase which reduces the formation of cAMP [41]. However, NPY also seems to inhibit insulin secretion through a mechanism distal to the formation of cAMP, because insulin secretion stimulated by a cAMP analogue can also be inhibited by NPY [1]. Although the physiologic importance of NPY for islet function remains to be established, there is some evidence that endogenous islet NPY tonically restrains insulin secretion. Thus, immunoneutralization of NPY in isolated islets leads to increased insulin secretion [42], and mice with genetic deletion of NPY have increased insulin secretion [43].

**Sympathetic effects and mediation**

Sympathetic nerves inhibit insulin and stimulate glucagon secretion. Their activation may help mediate stress-induced changes in islet hormone secretion including the glucagon counterregulatory response to hypoglycemia. These effects may be mediated by the combination of the classical neurotransmitter noradrenaline and the sympathetic neuropeptides (galanin and NPY). All three neurotransmitters can inhibit insulin and stimulate glucagon secretion, thus mimicking the effects of sympathetic activation on islet hormone secretion. In addition, galanin and NPY meet several other of the criteria needed to be classified as a physiologic neurotransmitter (see [30] and Tables 9.1 and 9.3).

**Islet sensory neuropeptides**

Islets are innervated by sensory, as well as autonomic, nerves. Some sensory nerves, usually the pain fibers, contain the

---

**Table 9.3** Application of the criteria from Table 9.1 for neurotransmitters in sympathetic islet nerves to mediate the inhibition of insulin secretion and stimulation of glucagon secretion which are effects achieved by activation of the sympathetic nerves

<table>
<thead>
<tr>
<th>Criterion 1</th>
<th>Criterion 2</th>
<th>Criterion 3</th>
<th>Criterion 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inhibits insulin secretion</td>
<td>Stimulates glucagon secretion</td>
<td>Localized to islet sympathetic nerves</td>
<td>Released from pancreas by sympathetic nerve activation</td>
</tr>
<tr>
<td>Galanin</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>NPY</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes*</td>
</tr>
</tbody>
</table>

*Also localized to parasympathetic nerves.
neuropeptides calcitonin gene-related peptide (CGRP) and substance P [1,44,45], which allows their localization by immunocytochemistry. These sensory islet nerves innervate mainly the peripheral portions of islets. They travel with the sympathetic nerves from the pancreas, reaching the spinal cord by way of the splanchnic nerves. The neurotoxin capsain causes deafferentation of small unmyelinated C-fibers and a substantial reduction of the number of islet CGRP nerves and a more moderate reduction of the number of islet substance P nerves [44]. Such chronic sensory denervation studies have shown that insulin secretion and glucose elimination are both increased in mice [46] and in Zucker diabetic fatty rats [47], suggesting that these sensory nerves tonically inhibit insulin secretion.

**Calcitonin gene-related polypeptide (CGRP)**

CGRP was originally described in a medullary thyroid carcinoma cell line, where it was found to be encoded in the same gene as calcitonin [45,48]. This gene, which is called the calcitonin complex gene A (CALCA) encodes for two different mRNAs, one that is translated to preprocalcitonin and the other translated to the 128-amino acid preproCGRP, which is further processed to α-CGRP. Later, another gene was described, named calcitonin complex gene B (CALCB), which encodes for β-CGRP. The α and β CGRPs are both 37-amino acid peptides and have a high degree of homology with each other and with islet amyloid polypeptide (IAPP) and adrenomedullin [48]. CGRP is localized both to the central and peripheral nervous systems. In the latter, it is largely found in sensory nerves. In the pancreas, CGRP nerves are scattered throughout the parenchyma, with particular density along small blood vessels and within islets where neurotransmitter release sites are in close apposition to the endocrine cells [45], suggesting an islet action. Exogenous administration of CGRP inhibits insulin secretion and in isolated islets, this inhibition is accompanied by reduction of the formation of cAMP which is prevented by pertussis toxin and therefore most likely mediated by an inhibitory Gi-protein [45].

CGRP, as well as IAPP, calcitonin, and adrenomedullin, can act through either the calcitonin receptor-like receptor (CRLR) or the calcitonin receptor (CTR). The affinity of these receptors for the different ligands is regulated: three different receptor activity-modifying proteins (RAMPs) govern the specificity of the receptors for the different ligands: RAMP1-transported CRLR is a CGRP receptor, whereas RAMP2- or RAMP3-transported CRLR is an adrenomedullin receptor. IAPP binding is more related to RAMP-transported CTR [49]. Of particular interest for islet physiology is the demonstration that CRLR and the RAMPs are expressed in β cells [49], suggesting that the regulation of RAMP1 expression and function governs the islet influence of CGRP. In conclusion, CGRP fulfills some, but not all, criteria for a sensory inhibitory islet neurotransmitter.

**Substance P**

Substance P was discovered in 1931 and is today known to be within the tachykinin family of peptides [50]. It consists of 11 amino acids and is ubiquitously distributed in sensory neurons with its main function involved in pain perception [50]. Immunohistochemistry of mouse and pig pancreas has demonstrated that substance P is expressed in islet sensory nerves [44,51]. It may therefore, together with CGRP, be involved in the regulation of islet function. This hypothesis is supported by the effect of acute capsain administration, which releases substance P from the pancreas [51]. The effects of substance P on insulin secretion remain, however, controversial. On one hand, substance P inhibits insulin secretion in mice and rats [52–54], whereas substance P stimulates insulin secretion in the pig pancreas [51]. There is more agreement regarding glucagon secretion: substance P stimulates glucagon secretion in both rats and pigs [51,52,55]. The main and preferred receptor for substance P is the neurokinin 1 receptor (NK1 receptor), which is a G protein-coupled receptor [56]. This receptor is abundantly expressed on α cells, but it has not been convincingly demonstrated to be expressed on β cells [57]. Therefore, although substance P is expressed in islet sensory nerves, controversy persists regarding its physiologic significance for islet hormone secretion.

**Summary of present-day knowledge**

We know today that the pancreatic islets are innervated by parasymathetic, sympathetic, and sensory nerves. The parasymathetic nerves stimulate insulin and glucagon secretion whereas sympathetic nerves inhibit insulin secretion and stimulate glucagon secretion; the net influence of the sensory nerves is not yet clearly established. We also know that besides the classical neurotransmitters, noradrenaline and acetylcholine, the islet autonomic nerves also release neuropeptides which can also influence insulin and glucagon secretion. PACAP, VIP, and GRP are released from parasymathetic nerves and can stimulate both insulin and glucagon secretion. Galanin and NPY are released from islet sympathetic nerves, can inhibit insulin secretion and stimulate glucagon secretion. CGRP and substance P can be released from islet sensory nerves with mixed effects on insulin and glucagon secretion. Several important questions remain to be answered to fully understand the role of the autonomic nerves in general, and their neuropeptides in particular, in the control of islet function.
Questions for future research

What are the physiologic roles of the islet autonomic nerves?

Although we know the effects on islet hormone secretion of activating autonomic nerves, there is less known about the physiologic conditions that activate these nerves and therefore their physiologic role. It is known that islet parasympathetic nerves are activated in anticipation of meals and early during the ingestive phase of a meal. Thus, they likely contribute to the release of insulin at this very early stage, the so-called cephalic phase of insulin secretion. There is also evidence for a parasympathetic contribution to insulin secretion even after glucose absorption. The contribution of autonomic nerves to meal-induced insulin secretion has been verified in an interventional study in healthy humans in which subjects were served a test meal together with an intravenous infusion of trimethaphan. This substance interrupts ganglionic transmission and therefore simulates autonomic denervation, at least to the extent possible in humans. The study showed that the early (10 min) increase in insulin was reduced by trimethaphan [3]. Thus, the likely physiologic role of islet parasympathetic nerves is to initiate and amplify the early insulin response to meals.

It is also known that pancreatic sympathetic nerves are activated during hypoglycemia. Since activation of these nerves stimulates glucagon secretion, it is likely that they contribute to glucagon counterregulation during hypoglycemia. This contribution was demonstrated in humans by inducing complete ganglionic blockade by trimethaphan during a hyperinsulinemic hypoglycemic clamp. It was found that the glucagon response was severely suppressed, which suggests that autonomic activation makes a major contribution to the glucagon response to hypoglycemia [58]. This autonomic stimulation of glucagon secretion is likely amplified by the direct action of low glucose to stimulate glucagon secretion from the α cells by releasing the α cell from the inhibitory effect the β cell has on it. In addition to these two relatively clear examples of autonomic contributions to islet hormone responses, there are likely others that have not been as carefully defined. In particular, the potential contribution of the sensory nerves needs to be examined in more detail.

What is the contribution of neuropeptides?

The classical autonomic neurotransmitters (acetylcholine for parasympathetic nerves and noradrenaline for sympathetic nerves) mimic many of the effects seen by direct activation of the nerves. However, the neuropeptides may also contribute to the effect of nerve activation, and Table 9.1 shows the criteria needed for acceptance of a neuropeptide as a physiologic neurotransmitter involved in the control of islet function. In Table 9.2 and Table 9.3, these criteria have been applied for the neuropeptides known to be present in parasympathetic and sympathetic nerves, respectively. Although some neuropeptides meet all these criteria, others need more study, particularly in humans. More generally, there is need for further study of the detailed regulation of small vesicles release (which contain the classical neurotransmitters) versus large dense core vesicles release (which contain the neuropeptides). In particular, do dissociated release mechanisms exist for these two types of vesicles and, if so, are they relevant for the regulation of islet hormone secretion?

Do the autonomic nerves contribute to development of type 2 diabetes?

Insulin resistance is a common feature of type 2 diabetes. Under normal circumstances, a compensatory increase in insulin secretion occurs thereby ensuring the maintenance of normal glucose tolerance. If this β-cell adaptation is insufficient for the degree of insulin resistance, glucose intolerance and eventually type 2 diabetes develop. Several signals may mediate the appropriate increase of insulin release, including circulating glucose and lipids. However, mediation by parasympathetic nerves may also contribute. First, people with insulin resistance have increased circulating levels of pancreatic polypeptide (PP) [59,60], which is a marker for the parasympathetic nerve activity. Second, the hyperinsulinemia of obese rats and mice is reduced after vagotomy [1] and the hyperinsulinemia induced by experimental insulin resistance in humans is reduced by ganglionic blockade [61]. Such studies provide proof of principle that the parasympathetic input to the islet is activated in states of insulin resistance and could, therefore, help prevent the failing insulin secretion in type 2 diabetes. However, further studies are needed to determine the contribution of defects in this parasympathetic pathway to the development of glucose intolerance and type 2 diabetes.

Is islet neurotransmission a therapeutic target for type 2 diabetes?

Due to the potent effects of the autonomic neurotransmitters on insulin secretion, it has been suggested that targeting their neurotransmitter receptors may be a potential therapy for type 2 diabetes. This has been of particular interest in relation to the parasympathetic receptors, since their activation increases insulin secretion. For example, activation of the M3 receptors [5] as well as of the PACAP [62,63], VPAC1 [64,65], and VPAC2 receptors [66] induces insulin secretion. A drawback is, however, that both PACAP and VIP also stimulate glucagon secretion, which would be predicted to aggravate hyperglycemia in type 2 diabetes; therefore this approach has not yet been approved for preclinical trials in humans.

Since the sympathetic neurotransmitters inhibit insulin secretion, interrupting their signals may be a therapeutic target, as has been suggested for α2-adrenoceptors [67]. Finally, the islet sensory nerves have also been proposed as a target for therapy, since sensory deafferentation improves diabetes in ZDF rats [47]. Again, both approaches need to be explored in more detail in preclinical human studies.
Acknowledgments

The author expresses his sincere gratitude to Dr. Gerald J Taborsky, Jr., Veterans Affairs Puget Sound Health Care System, University of Washington, Seattle, WA, USA, for very constructive criticism in the preparation of this chapter and continuous long-term discussions. The work by the author has been supported by grants from the Swedish Research Council (Grant No. 6834); the Faculty of Medicine, Lund University and Region Skåne.

References

3. Ahrén B, Holst JJ: The cephalic insulin response to meal ingestion in humans is dependent both on cholinergic and noncholinergic mechanisms and is important for postprandial glycemia. Diabetes 2001;50:1030–1038.


61 Ahrén B: Evidence that autonomic mechanisms contribute to the adaptive increase in insulin secretion during dexamethasone-induced insulin resistance in humans. Diabetologia 2008;51:1018–1024.

CHAPTER 10
Biosynthesis, secretion, and action of glucagon

Pierre J. Lefèbvre
University of Liège, Liège, Belgium

Key points
- The primary structure of glucagon has been remarkably conserved through evolution.
- The glucagon gene encodes for a preprohormone that contains not only glucagon, but several other peptides including GLP-1 and GLP-2.
- Glucagon acts through binding to specific receptors located at the target cell plasma membrane.
- The hepatocyte is the major target of glucagon.
- The control of glucagon secretion is multifactorial.
- Glucagon is the first line of defense against hypoglycemia.
- Glucagon plays a major role in starvation, physical exercise, stress, and the adaptation to extrauterine life.
- Plasma glucagon levels are increased in most forms of diabetes.
- Diabetes should be seen as a paracrinopathy of the islets of Langerhans in which intra-islet insulin deficiency results in excessive glucagon release from the neighboring α cells and the resulting hyperglucagonemia being a critical factor in the pathophysiology of diabetes [3,4]. A comprehensive bibliography on glucagon up to 1996 can be found in the three volumes of the Handbook of Experimental Pharmacology edited by Lefèbvre [5,6] and in the proceedings of a recently held conference [7].

Introduction

Glucagon, discovered in 1923 as a contaminant of early insulin preparations, has long been a neglected hormone.

Glucagon was among the very first polypeptide hormones to be isolated, purified, sequenced, and synthesized. Thanks to the pioneering work of Unger, it was the first polypeptide hormone to become measurable by radioimmunoassay, almost one year before insulin [1,2]. It has served two Nobel Prize winners as a unique tool, which permitted Sutherland and his associates to discover cyclic AMP (cAMP), and Rodbell and his coworkers to discover the role of G-proteins in cell-membrane receptors. Subsequently, the nucleotide sequence of the glucagon gene has been determined and the structure of the human glucagon precursor (preproglucagon) has been deduced. This discovery has been fundamental for clarifying the relationships of glucagon itself with various other peptides derived from the same common precursor, and originating from both the pancreas and the gut. Recent observations suggest that diabetes should be seen as a paracrinopathy of the islets of Langerhans in which intra-islet insulin deficiency results in excessive glucagon release from the neighboring α cells and the resulting hyperglucagonemia being a critical factor in the pathophysiology of diabetes [3,4]. A comprehensive bibliography on glucagon up to 1996 can be found in the three volumes of the Handbook of Experimental Pharmacology edited by Lefèbvre [5,6] and in the proceedings of a recently held conference [7].

Amino acid composition, extraction, synthesis, and biosynthesis

The amino acid sequence of glucagon isolated from the pancreas of pigs, cattle, and humans is identical. Guinea-pig glucagon is different from all other glucagons that have been isolated, while avian glucagons differ from the predominant mammalian glucagon only by a few conservative replacements. This extreme conservation of primary structure exhibited by glucagon has been considered “at least unusual and at most extraordinary.” In fact, it has now been recognized that glucagon belongs to what has been called the pituitary adenylate cyclase-activating polypeptide (PACAP)/glucagon superfamily that includes nine hormones that are related by structure, distribution (especially the brain and the gut), function (often by activation of cAMP), and receptors (a subset of seven transmembrane receptors) [8]. The nine hormones include glucagon, GLP-1 (glucagon-like peptide 1), GLP-2 (glucagon-like peptide 2), GIP (gastric inhibitory polypeptide), GRF (growth-hormone-releasing factor), PHM (peptide histidine methionine), PACAP, secretin, and...
Biosynthesis, secretion, and action of glucagon

VIP (vasoactive intestinal peptide). The origin of the ancestral superfamily members is at least as old as the invertebrates. The most ancient and tightly conserved members are PACAP and glucagon. Evidence to date suggests that the superfamily began with a gene or exon duplication and then continued to diverge with some gene duplications in vertebrates [8]. Mammalian glucagon contains 29 amino acid residues and has a molecular mass of 3485 Da (Figure 10.1). Glucagon can be synthesized in vitro by either classic solution synthesis or solid-phase synthesis; in both cases, two substrategies can be used—either fragment assembly or stepwise assembly. All procedures have given highly purified materials that are homogeneous and indistinguishable from natural glucagon by a range of sensitive analytic methods.

The glucagon gene consists of six exons and five introns spanning 10 kb; the human glucagon gene is located on chromosome 2 [9]. It encodes for a preprohormone of 180 amino acids that contains not only glucagon but several other peptides including two glucagon-like peptides whose amino acid structures are distinct from, but closely resemble, that of glucagon and other members of the glucagon superfamily of peptides. The glucagon gene is expressed in the α cells of the islets of Langerhans, the intestinal L cells, and in some parts of the brain. While the functional relevance of glucagon gene expression in the brain is still poorly understood, a clear image has emerged regarding the processing of proglucagon into several bioactive peptides in the pancreas and in the gut (Figure 10.2).

*In the pancreas* glucagon is the predominant peptide produced, together with glucagon-related polypeptide (GRPP), while the glucagon-like peptides remain in an incompletely processed prohormone fragment. *In the gut* two glucagon-like peptides (GLP-1 and GLP-2) are produced, while glucagon remains, in part, as a prohormone fragment, glicentin. However, glicentin can be further processed to oxyntomodulin and GRPP. Of these peptides of the glucagon superfamily, oxyntomodulin and possibly glicentin are implicated in the physiologic negative control of gastric acid secretion.

Glucagon-like peptide GLP-1(7-37), a peptide of 31 amino acids, or the equally potent isopeptide GLP-1(7-36), an amide of 30 amino acids, has major insulinotropic action on pancreatic β cells [10,11]. This peptide binds to specific receptors on islet β cells, stimulating cAMP formation, insulin release, proinsulin gene transcription, and proinsulin biosynthesis, all in a glucose-dependent manner. Interestingly, GLP-1 is a strong inhibitor of glucagon secretion. GLP-1 also exerts numerous extrapancreatic actions, including inhibition of food intake.

---

**Figure 10.1** The primary structure of human–porcine glucagon.

**Figure 10.2** Differential processing of proglucagon in the pancreas and in the gut. In the pancreas, the processing in the α cells is done by the prohormone convertase 2 and the main peptides released are glucagon, GRPP and a so-called “major proglucagon fragment.” In the intestinal L cells (as in the brain) the processing is done by the prohormone convertase 1/3 and the main peptides released are GLP-1, GLP-2, glicentin, and oxyntomodulin. IP-1 and IP-2 are intervening peptides.
intake, promotion of satiety, cardioprotection, vasodilation, and possibly beneficial effects on endothelial function and inflammation [12]. GLP-1 is rapidly degraded by the enzyme dipeptidyl-peptidase-4 (DPP-4). As reviewed in Chapter 48, GLP-1 analogues, such as exenatide and liraglutide, and DPP-4 inhibitors, such as vildagliptin, sitagliptin, saxagliptin, linagliptin, and alogliptin are now used in the treatment of type 2 diabetes.

GLP-2 is recognized as a major regulator of intestinal growth and function. Its secretion from the intestinal L cells is essentially stimulated by nutrient intake. It has specific trophic effects on the gut which appear to be mediated by stimulation of mucosal cell proliferation and inhibition of apoptosis and proteolysis [13]. Additional effects of GLP-2 on the digestive tract include stimulation of enteroocyte glucose transport and GLUT-2 expression, increased nutrient absorption, reduction of intestinal permeability, stimulation of intestinal blood flow, relaxation of intestinal smooth muscle, and inhibition of gastric emptying and gastric acid secretion [14,15]. Further effects of GLP-2 include stimulation of glucagon release from the islet α cells and modulation of islet adaptation to metabolic stress. GLP-2 has been reported to reduce bone resorption and to significantly increase bone mineral density in postmenopausal women. Preliminary studies suggest a great therapeutic potential of GLP-2 in total parenteral nutrition, short bowel syndrome following major intestinal resection, nonsteroidal drug-induced enteritis, inflammatory bowel disease, ischemic bowel, and so forth. Recently, teraglutide, a degradation-resistant GLP-2 analogue was approved for the treatment of short-bowel syndrome.

Both glucagon and oxyntomodulin are further processed into N-terminal and C-terminal fragments by cleavage at a dibasic site (Arg17-Arg18) [16,17]. The C-terminal fragments are of particular interest: glucagon-(19-29) modulates the plasma membrane calcium flow in the nanomolar range, whereas oxyntomodulin-(19-37) inhibits gastric acid secretion, as does oxyntomodulin itself. Further studies have shown that glucagon processing to glucagon-(19-29) (mini-glucagon) is probably essential for the positive inotropic effect of glucagon on heart contraction. Thus, the concept has emerged that glucagon and oxyntomodulin are first released into the blood and then processed at the level of their respective targets into the corresponding biologically active C-terminal fragments. Interestingly, mini-glucagon inhibits insulin secretion at picomolar concentrations. It has been proposed that mini-glucagon acts as a local inhibitory regulator of insulin release by turning off the main external calcium source for islet β cells via a specific receptor linked to ion channels that control cell polarity [18,19].

The tissue-specific liberation of proglucagon is controlled by cell-specific expression of enzymes known as prohormone convertases (PC). A major defect in the processing of proglucagon to mature pancreatic glucagon is seen in the PC2 knockout mouse. These animals exhibit mild hypoglycemia and hyperplasia of the α cells in the islets of Langerhans [20,21]. These abnormalities are corrected by glucagon replacement via a micro-osmotic pump [22]. Numerous studies have recently investigated the transcriptional regulation of α-cell differentiation and that of the proglucagon gene [7,23–25]. The differentiation of α cells and the maintenance of α-cell function are influenced at several stages during development and in the maturing islet. Several transcriptional factors such as neurogenin 3 (Ngn3), pancreatic duodenal homeobox 1 (Pdx-1) and regulatory factor X6 (Rfx6) are crucial in the determination of the endocrine cell fate. Other transcription factors such as aristaless-related homeobox (Arx) and forkhead box A2 (Foxa2) are implicated in the initial or terminal differentiation of α cells. On the other hand, proglucagon transcription, and therefore the maintenance of α-cell function, is regulated by several factors, including forkhead box A1 (Foxa1), paired box 6 (Pax6), brain 4 (Brn4), and islet-1 (isl-1) [23]. Interestingly, recent studies have shown that "endocrine cell reprogramming" is possible with conversion of pancreatic α cells into cells displaying a β-cell phenotype, thus possibly offering a totally new avenue for the treatment (or cure?) of diabetes [24,25].

**Physiologic effects of glucagon**

Glucagon acts through binding to specific receptors located at the target cell plasma membrane. The major common effect of glucagon is to activate adenylate cyclase and to increase the intracellular production of cAMP. There is considerable evidence that binding of glucagon to its receptor activates an intermediate transduction process that involves the participation of guanosine triphosphate (GTP), divalent cations, and adenosine (or other similar natural substances). The glucagon receptor has been characterized; it is a 62-kDa glycoprotein that contains at least four N-linked oligosaccharide chains and intramolecular disulfide bonds. The use of various mutated glucagon receptors has resulted in the demonstration that a five-amino acid domain within the membrane proximal portion of the COOH-terminal tail is required for cell surface expression; that most of the COOH-terminal tail is not required for glucagon binding, cAMP accumulation, or Ca2+ mobilization; and that phosphorylation of the COOH-terminal tail is required for glucagon-stimulated receptor internalization [26]. Further, positively charged residues at positions 12, 17, and 18 of the glucagon molecule contribute strongly to the stabilization of the binding interaction with the glucagon receptor that leads to maximal biologic potency [27].

Elimination of signaling through the glucagon receptor has been investigated in depth in mice [28]. Glucagon receptor knockout (GgcR/-) mice are viable but exhibit a number of striking phenotypes including an expected mild hypoglycemia and improved glucose tolerance. Other consequences include increase in total pancreatic weight, marked α-cell hyperplasia and extreme elevations of circulating plasma levels of glucagon and GLP-1. These mice also exhibit multiple defects in islet cell phenotypes implying a complex, and still poorly understood,
role of glucagon in islet development. Another striking observation is an increased susceptibility of GcgR-/- hepatocytes to apoptotic injury, suggesting that glucagon signaling is essential for cell survival. Furthermore, GcgR-/- hepatocytes exhibit profound defects in lipid oxidation and accumulate excessive lipid during fasting. Last, GcgR-/- mice exhibit reduced fetal weight, increased fetal demise at the end of gestation and abnormalities in the placenta suggesting that glucagon signaling is essential in fetal and placentual development. The most striking effect of elimination of glucagon signaling is the lack of diabetes after alloxan or streptozotocin destruction of islet β cells in GcgR-/- mice [29], as discussed later.

The hepatocyte is a major target cell of glucagon. The main effect of glucagon on the liver is to increase glucose output, an effect that results from inhibition of glycogen synthesis and stimulation of liver glycogenolysis and gluconeogenesis. Basal glucagon is important in maintaining hepatic glucose production during a prolonged fast and gluconeogenesis and glycogenolysis are equally sensitive to stimulation by glucagon in vivo [30,31]. Studies using mass isotopomers suggest that low doses of glucagon stimulate only nongluconeogenic glucose release while higher doses stimulate both the gluconeogenic and nongluconeogenic pathways [32]. Interestingly, hyperglucagonemia stimulated endogenous glucose production during fructose infusion, but this effect was not secondary to stimulation of gluconeogenesis but rather to channeling of glucose-6-phosphate toward systemic release [33,34]. There is ample evidence that most of these effects are mediated by cAMP, but the possibility has been raised that part of the glycogenolytic effect of glucagon may occur by a cAMP-independent mechanism. In vitro studies have shown that pulsatile delivery of glucagon is more efficient than continuous exposure to stimulate hepatic glucose production. Similarly, pulsatile delivery of glucagon in humans has greater effects in stimulating endogenous glucose production than continuous infusion. Furthermore, when both insulin and glucagon are delivered intermittently and out of phase, the greater effect of glucagon in stimulating glucose production prevails over the greater effect of insulin in inhibiting this parameter [35]. Another major effect of glucagon on the liver is to stimulate ketogenesis, which depends upon both the flux of free fatty acids (FFA) into the liver and the pathway status of this organ, which is influenced in a crucial manner by the glucagon/insulin ratio in the blood perfusing the liver [36]. Further, a high glucagon/insulin ratio increases the intracellular level of cAMP, reduces glycolysis and acetyl-CoA carboxylase activity, and reduces the intracellular concentration of malonyl-CoA. This fall in malonyl-CoA brings fatty acid synthesis to a halt and causes derepression of the enzyme carnitine acyltransferase such that incoming fatty acids (made abundant through stimulation of lipolysis) are efficiently converted into the ketone bodies acetocacetate and 3-hydroxybutyrate.

The effects of glucagon on the adipocyte markedly depend upon the species considered. Although glucagon is a potent lipolytic hormone in birds and in rodents, its effects on the human adipose cell have long been disputed. Recent investigations have shown that glucagon is indeed strongly lipolytic in the human adipocyte in vitro, but that this effect is difficult to demonstrate using incubation of adipocytes or adipose tissue pieces because glucagon is rapidly destroyed by a proteolytic activity associated with those cells. When perfusion techniques are used, the lipolytic effect of glucagon on human adipocytes can easily be demonstrated. In humans in the presence of somatostatin-induced insulin deficiency, pulsatile glucagon exerts greater effects than its continuous delivery not only on blood glucose, as mentioned earlier, but also on plasma FFA, glycerol, and 3-hydroxybutyrate levels [37]. Interestingly, in the older population the lipolytic and ketogenic, but not the hyperglycemic, responses to pulsatile glucagon are significantly reduced [37]. In healthy volunteers, moderate hyperglucagonemia undoubtedly stimulates the rate of appearance in the plasma of both glycerol and FFA [38], while, in contrast, microdialysis studies failed to demonstrate a lipolytic effect of glucagon [39,40]. It has been recently suggested that the lipolytic of glucagon in humans is mediated by FGF-21 (fibroblast growth factor 21) [41].

Other effects of glucagon

Other metabolic effects of glucagon include modification of the circulatory pattern of plasma amino acids (partly due to the stimulation of gluconeogenesis), a reduction in circulating levels of cholesterol and triglycerides, and a stimulation of fibrinogen formation. Glucagon also stimulates insulin release, but the physiologic character of this effect is questionable [42]. It has a major role, together with insulin, in liver regeneration [43]. Under certain circumstances, glucagon increases renal blood flow and glomerular filtration rate, and promotes renal loss of sodium and other ions [44]. At pharmacologic doses, glucagon stimulates adrenal catecholamine release, an effect that has been used for the diagnosis of pheochromocytoma. Combining glucagon stimulation and clonidine suppression testing has given a sensitivity of 100% and a specificity of 79% for the diagnosis of pheochromocytoma [45]. Glucagon also exerts positive inotropic and chronotropic effects on the heart [46], effects that may be useful, for instance, in treating the cardio depressive manifestations of poisoning by β-receptor blocking agents. Glucagon and several of its analogues, like glucagon-(1-21), which is devoid of metabolic effects, exert a potent smooth-muscle spasmytic action, sometimes used for various diagnostic procedures or for therapeutic applications like bronchospasm [47].

Control of glucagon release

There is considerable evidence that the control of glucagon secretion is multifactorial and involves direct effects of
nutrients on α-cell stimulus-secretion coupling as well as paracrine regulation by insulin, somatostatin and, possibly, other mediators such as zinc, γ-amino-butyric acid (GABA) or glutamate [48,49]. Glucagon secretion is also regulated by circulating hormones and the autonomic nervous system [50,51]. Table 10.1 lists the factors and conditions demonstrated as stimulators of glucagon secretion. The main physiologic or pathophysiologic stimulators of glucagon release are hypoglycemia (insulin-induced, associated with starvation or intense muscular exercise), hyperaminoacidemia (the rise in plasma glucagon levels after a balanced meal is probably due mainly to amino acid-induced glucagon release), stimulation of the adrenergic system (stress, exercise, and possibly hypoglycemia), and stimulation of the vagal system, which together with hormones like GIP and CCK-PZ probably participate in the mixed meal-induced glucagon rise.

The factors and conditions associated with inhibition of glucagon release are listed in Table 10.2. The main physiologic inhibitors of glucagon release are probably hyperglycemia and hyperinsulinemia (in a glucose-rich or carbohydrate-rich meal) and high circulating levels of FFA. It has been suggested that glucokinase may serve as a metabolic glucose sensor in pancreatic α cells, and hence constitute a mechanism for direct regulation of glucagon release by extracellular glucose. Intra-islet insulin, glucagon, and somatostatin release have been shown to be interrelated [52]. In such paracrine mechanisms, further data suggest that the oscillatory pattern of islet hormone release may be particularly important [53,54]. Using isolated perifused human islets, it has been shown that glucose generates coincident insulin and somatostatin pulses and clear antisynchronous glucagon pulses [55]. The periodicity of these pulses is 7 to 8 min. The fact that these pulses occur in isolated islets demonstrates that their origin is the islets themselves and independent of external metabolic, hormonal, or neuronal signals. The nature of the intra-islet signal(s) coordinating the secretion of the various endocrine cells of the islets of Langerhans is still the subject of intense investigation [56–61].

### Some aspects of glucagon physiology and pathophysiology

#### Glucagon as a counterregulatory hormone

Numerous studies have shown that the liver is the main site at which moment-to-moment control of glucose homeostasis takes place and that in normal humans glucagon is the major glucose counterregulatory hormone. By antagonizing the suppressive effects of insulin on glucose production and by stimulating glucose production when appropriate, glucagon not only defends the organism against hypoglycemia, but also restores normoglycemia if hypoglycemia occurs. Perturbation of the mechanisms controlling hypoglycemia-induced glucagon release...
release in some diabetic patients markedly increases the risk of severe hypoglycemia in these subjects. Other hormones, such as epinephrine (acutely) and growth hormone and cortisol (more slowly), participate in the counterregulation of the effects of insulin, but careful clinical observations suggest that indeed glucagon is the first line of defense against hypoglycemia [62].

**Glucagon in exercise**
Glucagon levels increase progressively during prolonged exercise [63], during which blood glucose remains relatively constant thanks to a fine balance between muscle glucose uptake and liver glucose production. Although a rise in plasma glucagon does not appear to be essential for increased glucose production during exercise, the presence of glucagon does appear to be necessary.

**Glucagon in stress**
Hyperglucagonemia is a classic feature of stress [64]. It occurs mainly as a result of the β-adrenergic stimulation associated with stress and undoubtedly contributes to the hyperglycemia, which is a classical finding in this condition.

**Glucagon in starvation**
Starvation is accompanied by a decline in circulating insulin and a moderate rise in plasma glucagon [65]. The main effects of glucagon during starvation are at the liver, where it contributes to the maintenance of continuous liver glucose output (initially by stimulating glycogenolysis, and later by promoting gluconeogenesis) and the induction of ketogenesis. Whether glucagon contributes to the stimulation of adipose tissue lipolysis during starvation is still disputed.

**Glucagon and adaptation to extrauterine life**
A significant rise in plasma glucagon occurs soon after birth in all species investigated so far, which suggests that glucagon has a crucial role in neonatal glucose homeostasis [66]. Furthermore, an important role of glucagon in thermogenic regulation has been suggested.

**Glucagon and diabetes**
Plasma levels of glucagon have been found to be increased in all experimental and clinical forms of diabetes mellitus. This disturbance undoubtedly contributes to the hyperglycemia of the disease and excessive ketogenesis of diabetic coma. Numerous studies have shown that failure of glucagon suppression contributes to postprandial hyperglycemia in type 1 [67] and type 2 [68,69] diabetes. Impaired glucagon suppression contributes with impaired insulin release to the excessive blood glucose levels in early type 1 diabetes [70], in subjects with impaired glucose tolerance [71,72], and in patients with ketosis-prone atypical diabetes [73,74]. Morphologic studies have established that the main abnormality in the islet cell population of diabetes is a decrease in the β cells with a relative expansion of the α-cell mass [75].

The proposal of Unger and Orci to consider diabetes as a "paracrinopathy" of the islets of Langerhans [3] is based on the concept that the very high concentrations of insulin normally reached inside the stimulated islet exerts, directly or by proxy, a major inhibitory effect on glucagon secretion from the neighboring α cells. Conversely, a reduction in intra-islet insulin concentrations would permit glucagon release from the α cells. Disruption of this mechanism is proposed as a key factor in the pathophysiology of diabetes [76]. This concept is supported by recent data on the micro-anatomy of the islets of Langerhans [74,77]. In type 1 diabetes, α cells lack constant action of high insulin levels from juxtaposed β cells. Replacement with exogenous insulin subcutaneously injected does not approach the paracrine levels of insulin, except with high doses that “overinsulinate” the peripheral insulin targets, thereby promoting glycemic volatility [3]. In type 2 diabetes, the α-cell dysfunction may result from the failure of the juxtaposed β cells to secrete the first phase of insulin or from the loss of the intra-islet pulsatile secretion of insulin. Observations made in experimental diabetes in minipigs [78] and recently confirmed in human type 2 diabetes [79] are in support of the second mechanism.

Inhibition of glucagon secretion markedly improves experimental diabetes in rodents [80] and knockout of the glucagon receptor makes rodent models of insulin-dependent type 1 diabetes thrive without insulin [29]. The critical role of glucagon action in the liver in diabetes has been demonstrated by expressing glucagon receptors in livers of glucagon receptor-null (GcgR-/-) mice before and after β cell destruction by high doses of streptozotocin [81]. Wild-type mice developed fatal diabetic ketoacidosis after streptozotocin, whereas GcgR-/- mice remained clinically normal without hyperglycemia, impaired glucose tolerance, or hepatic glycogen depletion. Restoration of receptor expression using adenovirus containing the GcgR cDNA restored hepatic GcgR, phospho-cAMP response element binding protein (P-CREB) and phosphoenol-pyruvate carboxykinase, markers of glucagon action rose dramatically and severe hyperglycemia appeared. When GcgR mRNA spontaneously disappeared 7 days later, P-CREB declined and hyperglycemia disappeared. In this experimental setting, the metabolic manifestations of diabetes cannot occur without glucagon action, and once present, disappear promptly when glucagon action is abolished.

These observations strongly suggest that targeting the α cell and glucagon are innovative approaches in diabetes management. Compounds such as pramlintide and GLP-1 agonists reduce glucagon release and this is considered an important component of their antidiabetic action (see Chapter 48). On the other hand, numerous glucagon antagonists, either peptidic or nonpeptidic, have been indentified and some have entered clinical trials. However, marked inhibition of glucagon signaling may result in α-cell hyperplasia, increased mass of the pancreas, increased susceptibility to hepatosteatosis and hepatocellular injury, and an increased risk of hypoglycemia [82]. Further studies in normal and diabetic subjects should identify the
extent to which reduction of glucagon signaling produces a compelling therapeutic benefit without incurring a risk of adverse events.

**The glucagonoma syndrome**

The glucagonoma syndrome is a rare disorder associating necrolytic migratory erythema, cheilosis, usually mild diabetes mellitus, anemia, weight loss, venous thrombosis, and, frequently, neuropsychiatric symptoms [83].

**References**

64 Lindsey CA, Faloona GR, Unger RH: Plasma glucagon levels during rapid exsanguination with and without adrenergic blockade. *Diabetes* 1075;24:313–319.


CHAPTER 11
Incretin physiology in health and disease

Juris J. Meier and Michael A. Nauck
Division of Diabetes and Gastrointestinal Endocrinology, St. Josef-Hospital, Ruhr-University, Bochum, Germany

Key points
• The postprandial augmentation of incretin hormones by gastrointestinal hormones is called the incretin effect.
• In humans, gastric inhibitory polypeptide (GIP) and glucagon-like peptide 1 (GLP-1) are the major incretin hormones.
• GLP-1 stimulates insulin secretion, suppresses glucagon, delays gastric emptying, and inhibits food intake.
• GIP stimulates insulin secretion, but even slightly stimulates glucagon release.
• A deficit in incretin stimulation is characteristic of patients with type 2 diabetes. This is largely attributable to the reduced insulinotropic effect of GIP.
• The defects in the incretin system in patients with type 2 diabetes are likely a consequence rather than the cause of diabetes.

The incretin effect in health

The idea that gastrointestinal factors contribute to the control of postprandial glucose regulation dates back to the beginning of the twentieth century, when Moore and colleagues reported reductions in glucosuria after the oral administration of gut extracts in patients with juvenile diabetes [1]. Even though it is questionable whether these glucose-lowering effects were really attributable to the incretin activity of the extract (which is unlikely, because most gastrointestinal peptide hormones are inactivated by the gastric acid), this report can be considered as the first description of an incretin-like effect. The term “incretin” was coined in 1932 by La Barre to describe a putative substance from the upper gut mucosa, which lowers glucose without activating exocrine secretion [2,3]. At this early time period, various reports about glucose-lowering activities of gut extracts had appeared. However, probably owing to the variations in the purity of these extracts, the results of these experiments were quite heterogeneous, and the idea that gut hormones contribute to glucose regulation was not explored further until the 1960s [3].

The development of the first radioimmunoassay for insulin in 1962 by Berson and Yallow opened up a new era in metabolic research. Elrick and colleagues [4] and McIntyre and colleagues independently reported that oral administration of glucose led to significantly higher increments in insulin secretion than intravenous glucose administration [5], thereby suggesting the existence of gut-derived factors that augmented insulin secretion, that is, an “incretin effect” (Figure 11.1). While these initial studies had therefore already demonstrated some incretin activity in humans, it was still difficult to quantify the respective contribution of this effect to the overall postprandial rise in insulin levels, because the circulating plasma glucose concentrations after i.v. glucose injection significantly exceeded those after oral glucose administration, thereby confounding direct comparisons of the respective insulin concentrations. Therefore, Nauck and colleagues developed the “isoglycemic clamp technique” to quantify the contribution of the incretin effect [6] (Figure 11.2). In these experiments, glucose was first administered by mouth, and the plasma glucose rises were recorded at 5-min intervals. On a second day, glucose was administered intravenously, and the respective glucose infusion rates were varied to exactly copy the glucose concentration pattern of the first experimental day. By these means, it was possible to compare the rises in insulin and C-peptide concentrations at identical plasma glucose level [6]. The incretin effect was then calculated by the equation

\[
\text{Incretin effect} = \frac{(\text{AUC}_{\text{oral glucose}} - \text{AUC}_{\text{intravenous glucose}})}{\text{AUC}_{\text{oral glucose}}} \times 100\%.
\]
based on the integrated incremental responses (AUCs) of insulin or C-peptide concentrations or of insulin secretion rates derived by deconvolution of C-peptide concentration-time curves.

Based on these studies, it was estimated that the incretin effect contributed between 27.6% and 62.9% to the overall increments in C-peptide levels after oral glucose ingestion [6]. It was also demonstrated that the “size” (i.e., the percentage contribution) of the incretin effect increased with greater amounts of glucose being administered [6]. Because insulin is subject to considerable first-pass clearance by the liver, the C-peptide-based calculation was considered to be more accurate than the calculation using insulin concentrations. In fact, hepatic insulin clearance is significantly reduced by oral, but not by intravenous glucose administration, meaning that calculations based on insulin levels might lead to an over-estimation of the incretin effect.

While augmentation of insulin secretion is clearly the most prominent action of the incretin hormones, there also appears

---

**Figure 11.1** The incretin effect on insulin (a) and glucagon (b) secretion. Following oral administration of nutrients, islet hormone secretion is modulated by (A) the incretin hormones GIP and GLP-1, (B) the sympathetic and parasympathetic nervous system, and (C) the circulating substrates glucose, amino acids, and free fatty acids. Stimulatory actions have been denoted with blue color, inhibitory actions are indicated by red color. Source: Adapted from Creutzfeldt 1979 [98]. Reproduced with permission of Springer. (For a color version of this figure, please see color plate section.)
to be an effect on glucagon secretion (Figure 11.1). Thus, when glucagon levels after oral and isoglycemic intravenous glucose administration were directly compared, the suppression of glucagon levels was found to be greater after intravenous compared to oral glucose administration [7]. This finding is surprising given the glucagonostatic actions of GLP-1 [8]. However, it appears that these glucagon-lowering effects of GLP-1 are outweighed by the glucagonotropic actions of GIP and potentially GLP-2 [9]. Indeed, when GIP was exogenously co-infused along with GLP-1, the GLP-1-induced suppression of glucagon was completely abolished [10]. Therefore, the net incretin effect on glucagon levels appears to be a modest stimulation of α-cell secretion under normal conditions.
Although the quantification of the incretin effect based on the isoglycemic clamp technique has been accepted as a reference standard for most subsequent studies, a couple of methodological caveats need to be borne in mind. Thus, matching circulating plasma glucose concentrations with oral and intravenous glucose administration inevitably results in differences in the amount of glucose being administered on both occasions. Thus, typically $\sim 20–25$ g of i.v. glucose were required to reach isoglycemia with $50$ g of glucose ingested orally [6]. One might therefore argue that the study design could overestimate the incretin effect due to the different amounts of glucose being administered [11]. Along this line, another argument to consider is the fact that the oral glucose load engages pancreatic $\beta$ cells through various mechanisms (direct stimulation by circulating glucose, afferent neutral signals, and humoral stimulation by incretin hormones), whereas i.v. glucose represents a single $\beta$-cell stimulus. It seems conceivable that under conditions of a general impairment of $\beta$-cell mass or function the relative impairment in insulin secretion might be greater with the large combined stimulus (oral glucose) than with the smaller stimulus (i.v. glucose) [11]. It is also important to bear in mind that the estimation of the incretin effect based on isoglycemic clamp experiments is not necessarily representative of the postprandial situation, because the potential contributions of the fat and protein components of a typical mixed meal and other gastrointestinal hormones with insulinotropic effects released by these food components are not taken into account. In line with this, studies in humans have demonstrated significant incretin effect after oral fat or amino acid ingestion as well [12,13]. Finally, variations in the rate of gastric emptying may alter the patterns of insulin secretion after oral glucose ingestion, independent of incretin activity.

Other groups have therefore applied other techniques, such as a relative comparison of insulin secretion after oral glucose ingestion during fixed hyperglycemia at 8 and 10.5 mmol $L^{-1}$ in a single experiment [14]. However, such study design does not take into account the typical postprandial changes in plasma glucose concentrations.

Taken together, the quantification of the incretin effect is complex and can be potentially confounded by various factors. Despite various limitations, the isoglycemic clamp technique still seems to represent the most appropriate measure of the incretin effect.

**Secretion of incretin hormones**

The mucosa of the intestinal tract harbors a large number of endocrine cells that give rise to various peptide hormones. These include cholecystokinin, motilin, secretin, gastrin, gastric inhibitory polypeptide (GIP, also referred to as glucose-dependent insulinotropic polypeptide), glucagon-like peptide 1 (GLP-1), glucagon-like peptide 2, and peptide YY [15].
tyrosine tyrosine (PYY) are further intestinal proglucagon cleavage products [27]. GLP-1 is secreted in two distinct forms, an amidated form GLP-1 (7-36) amide, and a glycine-extended form GLP-1 (7-37). The amidated form appears to be more abundant [27]. The biologic activity of these forms has been found to be rather similar [16].

In humans, GLP-1 secretion is enhanced by glucose, triglycerides as well as (to a lesser degree) various amino acids in vivo [16]. The mechanisms of GLP-1 secretion have been further examined in an intestinal L-cell line, called GLUTag cells, as well as in primary rodent L-cell cultures [28]. In these experimental models, expression of the sulfonylurea receptor subunits Kir6.2 and SUR1 as well as glucokinase have been detected [29]. Furthermore, stimulation of GLP-1 release through SGLT1 transporters has been suggested [29]. There is also wide evidence for a role of G-protein coupled receptors (GPRs), especially GPR 40, GPR 120, and GPR 119, in GLP-1 secretion [29]. These receptors have been demonstrated to be of particular importance for fatty acid-induced GLP-1 secretion. Stimulation of GLP-1 secretion by bile acids has also been reported in different models. These effects have been suggested to be mediated through the G protein-coupled receptor TGR5, with subsequent rises in intracellular cAMP levels [30]

Various studies in rodent models have suggested that artificial sweeteners, such as saccharin, acesulfame potassium, d-tryptophan, and sucralose, also stimulate GLP-1 release by a newly discovered group of sweet taste receptors [31]. Human studies using the sweet taste receptor T1R2/T1R3 antagonist lactisole have confirmed a physiologic role of these receptors in GLP-1 secretion [32].

There might also be a paracrine inhibitory effect of locally produced somatostatin on L-cell secretion. Finally, recent studies in rodents have suggested a stimulation of GLP-1 release by interleukin-6 (IL-6) [33].

**Degradation and elimination of incretin hormones**

Following their intestinal secretion, the incretin hormones GIP and GLP-1 are subject to rapid proteolytic cleavage by the enzyme dipeptidyl-peptidase 4 (DPP-4) [34]. DPP-4 can be found throughout the vascular endothelium as well as circulating in the plasma [35]. Because endogenously secreted GLP-1 enters the intestinal capillary network first and is subsequently delivered into the portal venous circulation, only a minor proportion of GLP-1 reaches the systemic circulation in its intact form GLP-1 (7-36) amide, whereas ~85% of the total GLP-1 concentration in the peripheral circulation is found in the (largely) inactive form GLP-1 (9-36) amide [36]. Thus; ~75% of the secreted GLP-1 are inactivated already within the gut, and an additional 40–50% of the remaining GLP-1 undergo degradation during the first liver passage [36]. It has also been suggested that another circulating enzyme, neutral endopeptidase 22.11, contributes to the degradation of GLP-1 to a large extent (up to 50%) [37]. While proteolytic degradation is therefore apparently the most critical step in the inactivation of GLP-1, renal filtration is driving the final elimination of the intact peptide as well as the major metabolites [38]. The mechanisms of renal handling of GLP-1 seem to involve both glomerular filtration and tubular uptake and catabolism [38]. In line with this, human studies have revealed increased plasma concentrations of GLP-1 in patients with renal impairment [38]. The plasma half-life of GLP-1 was calculated as 2.3 ± 0.4 min for the intact peptide and 3.3 ± 0.4 min for the major metabolite GLP-1 (9-36) amide. The corresponding metabolic clearance rates were 2.42 ± 0.45 L min⁻¹ and 0.64 ± 0.16 L min⁻¹ for intact GLP-1 and the metabolite, respectively [38].

Like GLP-1, the intact peptide GIP (1-42) is primarily inactivated through proteolytic cleavage by DPP-4, yielding the major metabolite GIP (3-42), and both the intact peptide and the metabolite are subject to renal elimination [39]. The half-life of GIP is considerably longer than that of GLP-1 and has been calculated as 5.0 ± 1.2 min for the intact peptide and 22.4 ± 3.0 min for the metabolite. The respective metabolic clearance rates were 3.18 ± 0.62 L min⁻¹ and 1.56 ± 0.27 L min⁻¹ [38].

**Physiologic effects of incretin hormones**

**GLP-1**

The GLP-1 receptor has been detected on numerous organs and tissues. High levels of GLP-1 receptors have been found on pancreatic β-cells, δ-cells, endothelial cells, cardiomyocytes, the gastrointestinal tract, and the kidneys [40]. The actions of GLP-1 have been summarized in Figure 11.3. There is an ongoing debate as to the expression of GLP-1 receptors on pancreatic acinar cells and thyroid C-cells, with discrepant results being reported in different studies. Also, the expression of GLP-1 receptors of pancreatic α-cells has been controversially discussed. Whereas most studies have failed to detect such receptors on α-cells, others have reported small GLP-1 receptor numbers [40]. The discrepant results of different studies may be partly related to methodological problems with unequal specificities and binding characteristics of different commercially available GLP-1 receptor antibodies.

**Modulation of islet hormone secretion**

Intravenous administration of GLP-1 under fasting conditions leads to an immediate rise in insulin secretion [8,41]. These effects can be detected both at pharmacologic as well as physiologic plasma concentrations. In line with this, GLP-1 receptor knockout mice display significant reductions in insulin secretion [42]. Notably, GLP-1 stimulates insulin secretion only in the presence of hyperglycemia [43], and various experiments have demonstrated that the insulinoactive effect ceases once
normoglycemia has been reached [41]. Based on stepwise hypoglycemic clamp experiments, plasma glucose concentrations of more than \(\sim 65 \text{ mg dL}^{-1}\) will be necessary to allow for GLP-1 to augment insulin release [43].

The \(\beta\)-cell effects of GLP-1 have been shown to be mediated through binding to the GLP-1 receptor on the cell surface [44]. Intracellular signaling involves generation of cAMP via protein kinase A [45]. This favors closure of ATP-dependent potassium channels, opening of voltage-gated calcium channels and exocytosis of insulin granules at elevated glucose concentrations [45]. Coupling of these processes to glucokinase action explains the glucose-dependency of the GLP-1 effects. However, GLP-1 alone cannot close ATP-dependent potassium channels and initiate insulin secretion at low glucose concentrations.

In addition to the direct insulinotropic effect, GLP-1 has also been demonstrated to enhance insulin biosynthesis [46]. Accordingly, islet insulin content was found to be increased after incubation with GLP-1 under culture conditions [47]. These properties may explain the durability of GLP-1 action on insulin secretion. Finally, experiments in isolated \(\beta\) cells have suggested that GLP-1 treatment confers glucose-competence to previously quiescent \(\beta\) cells [48]. This means that GLP-1 may recruit additional, previously inactive, \(\beta\) cells to the insulin secretory machinery. As a clinical correlate to these cell culture experiments, improvement of glucose-responsiveness in patients with diabetes has been demonstrated in “glucose ramp” experiments [49].

Suppression of glucagon concentrations is the second major islet cell effect of GLP-1 [8,41]. Similar to the insulinotropic effect, the suppression of glucagon levels by GLP-1 is strictly glucose-dependent, and hypoglycemic clamp experiments did not reveal any alterations of hypoglycemia counterregulation during GLP-1 infusion [43]. The effect of endogenous GLP-1 is also believed to be relevant for the physiologic control of glucagon secretion, since experiments with i.v. administration of the GLP-1 receptor antagonist exendin (9-39) have demonstrated a significant increase in glucagon concentrations after oral glucose ingestion [50]. There is some controversy as to the mechanisms underlying the suppression of glucagon secretion by GLP-1, since most studies did not detect sufficient amounts of GLP-1 receptors on pancreatic \(\alpha\) cells. Inhibition of glucagon secretion secondary to the increase in insulin secretion cannot plausibly explain the effects either, because a similar reduction in glucagon levels has also been reported in patients with type 1 diabetes [51]. However, the finding that co-administration of a somatostatin receptor 5 antagonist abolishes the effects of GLP-1 on glucagon secretion in the isolated perfused pancreas suggests that the glucagonostatic effect of GLP-1 is secondarily mediated through an increase in somatostatin release from pancreatic \(\delta\) cells [52].

Finally, reduction of PP secretion has been demonstrated in response to GLP-1 administration. This effect is held to be mediated via an inhibition of vagal nervous activation [53].

**Gastrointestinal effects**

Inhibition of gastric emptying during intravenous infusion of GLP-1 has been demonstrated using various techniques both in healthy individuals and in patients with type 2 diabetes [54–56]. Unlike the effects on islet cell secretion, the deceleration of gastric emptying by GLP-1 is rather independent of plasma glucose concentrations, but exhibits a clear dose-response relationship [56]. More specifically, a prolongation of the lag period along with a dose-dependent inhibition of antral propagations and antroduodenal contractions has been reported during GLP-1 administration.
The finding that administration of the GLP-1 receptor antagonist exendin (9-39) increases gastric antral motility and reduces pyloric tone suggests a role for endogenous GLP-1 in the physiologic control of gastric emptying. These actions are believed to contribute to the so-called “ileal brake”, that is, the inhibition of gastric emptying by humoral and neural signals from the lower parts of the small intestine [57].

A transient reduction of exocrine pancreatic secretion has also been reported during GLP-1 administration [54]. However, it is difficult to judge whether these effects are due to a direct pancreatic effect of GLP-1 on the exocrine pancreas or rather secondary to the delay in gastric emptying leading to reduced secretion of gastrointestinal hormones that would in turn stimulate exocrine pancreatic secretion [58].

Furthermore, some earlier studies in humans have suggested a modest reduction in gastric acid secretion by intravenous GLP-1 infusion [59].

Mechanistically, the deceleration of gastric emptying by GLP-1 is believed to be mediated by an inhibition of vagal activation [53]. In support of this, co-administration of atropine in humans as well as afferent vagal denervation in rats have abolished the gastric effects of GLP-1 [60].

Because retardation of gastric emptying inevitably results in delayed and protracted glucose appearance in the systemic circulation, postprandial insulin concentrations are typically reduced during GLP-1 administration [55,56]. This observation has led to some controversy as to whether GLP-1 really fulfills the classical criteria of an incretin hormone (i.e., an augmentation of postprandial insulin secretion at typical plasma concentrations) [61]. The importance of the gastric emptying effect of GLP-1 also becomes apparent from experiments, in which the prokinetic drug erythromycin was co-administered along with GLP-1 in order to antagonize the deceleration of gastric emptying. Under those conditions, the GLP-1-induced reduction of postprandial glycemia was less pronounced than with GLP-1 alone, and the inhibition of insulin release was largely abolished [62].

Notably, the effect of GLP-1 on gastric motility appears to be subject to rapid tachyphylaxis and tends to wane with continued exposure to high GLP-1 plasma concentrations.

Central nervous effects
A role for GLP-1 in the central nervous regulation of food intake has been inferred from the high density of GLP-1 receptors in the hypothalamus [63]. Also, direct intracerebroventricular administration of GLP-1 has caused a significant reduction in food intake in rats, and these actions could be abolished by co-administration of the GLP-1 receptor antagonist exendin (9-39) [63]. These initial effects have been confirmed and extended by several other groups. Some studies have also reported a reduction of fluid intake by GLP-1 at pharmacologic concentrations [64]. Three potential sources of central nervous GLP-1 action have been suggested: (1) there is evidence for a local production of GLP-1 in the brainstem, from where the peptide is believed to be transported along axonal networks into the CNS [65]; (2) studies in humans have demonstrated that peripherally produced GLP-1 can access the brain through certain areas that lack a typical blood–brain barrier, such as the area postrema and the subfornical organs [66]; (3) receptors on peripheral nerve ends in the gut mucosa, the portal venous bed as well as the vagal nerve are held to project to certain areas within the hypothalamus to exert action on appetite regulation and glucose metabolism [36].

In humans with and without diabetes, significant reductions of appetite sensations and food intake have been demonstrated during the exogenous administration of supraphysiologic amounts of GLP-1 [67]. These effects were found to be dose-dependent, and significant effects on food intake have only been observed with relatively high doses [68]. It is therefore questionable whether the endogenous secretion of GLP-1 plays a major role for the postprandial induction of satiety.

Other potential GLP-1 effects of unknown physiologic relevance
Aside from the well-characterized action on islet hormone secretion, gastrointestinal motility, and appetite regulation, a number of potential extra-glycemic actions of GLP-1 have been suggested. Evidence for these actions has been derived primarily from experimental studies in rodents or in vitro models. The physiologic significance of these actions is therefore largely unclear.

Regulation of islet cell death and turnover
Induction of β-cell proliferation during incubation with GLP-1 was first observed in β-cell lines as well as in isolated rodent islets [69]. Subsequently, these results have been confirmed in a large number of rodent studies, in which both islet neogenesis and replication of existing β cells have been described after treatment with GLP-1 or GLP-1 receptor agonists [70]. It has also been suggested that GLP-1 treatment promotes transdifferentiation of exocrine ductal cells into insulin-secreting β cells and induces expression of Pdx-1 in the pancreas [71,72]. The fact that GLP-1 receptor knockout mice exhibit abnormalities in islet cell architecture indicates some physiologic relevance of these effects for normal islet formation [73]. However, due to the inaccessibility of the human pancreas for biopsy sampling, it is as yet unclear whether GLP-1 also promotes β-cell proliferation in humans to a measurable extent. Furthermore, the vast majority of rodent studies demonstrating induction of β-cell proliferation with GLP-1 have been carried out in neonatal or infant rodent models, and more recent studies in adult mice and rats have called into question the stimulation of β-cell proliferation with GLP-1 receptor agonists, even at pharmacologic concentrations [74]. There is also evidence that the proliferative capacity of human β cells is much lower compared to rodent islets. In addition to β-cell proliferation, effects on β-cell apoptosis have also been ascribed to GLP-1. In isolated human islets, apoptosis was significantly reduced during incubation with the peptide [47],
and these effects have been reproduced in various animal and in vitro models [69,75]. Mechanistically, the reduction of β-cell apoptosis by GLP-1 is held to be mediated through a cAMP- and PI3K-dependent signaling pathway. The relevance of these preclinical findings for human physiology cannot be judged with certainty at present.

**Cardiovascular effects of GLP-1**

A potential cardioprotective role for GLP-1 has been initially suggested based on the indirect effects on various cardiovascular risk factors, such as glycemia, blood pressure, hyperlipidemia, and obesity [76]. Interestingly, whereas numerous studies have consistently demonstrated significant reductions in blood pressure during therapeutic administration of various GLP-1 receptor agonists [77], initial experiments in rodents had demonstrated the opposite effect of GLP-1 to even raise blood pressure [78]. GLP-1 receptors have also been identified directly on cardiomyocytes as well as on vascular smooth muscle cells [79]. A number of experimental studies have examined the effects of GLP-1 in ischemia-reperfusion models. Most of these studies have reported significant reductions in myocardial necrosis along with preserved contractility after exogenous administration of pharmacologic doses of GLP-1 or GLP-1 receptor agonists [80]. Furthermore, improved myocardial glucose uptake has been reported in dogs, and studies in experimental models of cardiomyopathy [81] or in humans with chronic heart failure have suggested a positive inotropic effect of GLP-1. There is also evidence for enhanced flow-mediated vasodilation after GLP-1 administration in humans. In rodent models of atherosclerosis, GLP-1 receptor activation has been shown to directly prevent the formation of atherosclerotic lesions [82]. Mechanistically, these cardioprotective properties of GLP-1 have been largely attributed to binding of GLP-1 to cardiac receptors with subsequent activation of intracellular PI3-kinase-dependent pathways [82]. More recently, the finding that cardioprotective GLP-1 effects were also observed in mice lacking a GLP-1 receptor as well as after administration of the GLP-1 metabolite (9-36) amine, which does not bind to the typical GLP-1 receptor, have led to speculation about a second GLP-1 receptor in the cardiovascular system [82]. However, as yet no such receptor has been identified. In recent clinical trials examining the effects of pharmacologic doses of GLP-1 receptor agonists, increases in heart rate of 3–6 beats per min were reported [82]. Overall, a large number of studies have demonstrated effects of GLP-1 in the cardiovascular system at supraphysiologic plasma concentrations. The importance of such effects in normal physiology is still largely unclear.

**Neuroprotection**

Mice with a GLP-1 receptor knockout are characterized by a learning deficit, which was found to be reversible after gene transfer of the GLP-1 receptor into the hippocampus, thereby suggesting a role for GLP-1 in learning processes [83]. Various rodent studies have also demonstrated that neuronal cell damage can be partly prevented by the administration of GLP-1 or GLP-1 receptor agonists [83]. Finally, neuroprotective effects as well as enhanced neurogenesis have been reported in rodent models of Alzheimer’s disease, Parkinson’s disease, and Huntington’s disease. The physiologic relevance of these preclinical findings is difficult to judge at present [84]. Other effects of GLP-1 that have been suggested include enhanced diuresis and renal excretion of sodium chloride, increased skeletal muscle glucose uptake and enhanced liver glycogen storage, and enhanced bone formation [84].

**GIP**

Similar to GLP-1, GIP acts through specific receptors belonging to the seven transmembrane-domain family of receptors. These receptors are expressed in the endocrine pancreas, on adipocytes, and in the brain [85]. The actions of GIP have been summarized in Figure 11.3.

**Islet hormone secretion**

Intravenous administration of porcine or human GIP has led to a dose-dependent stimulation of insulin secretion in humans. These effects were found to be glucose-dependent, with negligible insulinotropic activity at normoglycemia. The physiologic importance of these effects has become apparent from experiments with immunoneutralization of circulating GIP or in GIP receptor knockout mice. Both ways to eliminate GIP signaling have led to glucose intolerance in the respective experimental animals [85]. Also, when the GIP plasma concentrations that are typically measured after meal ingestion were mimicked by exogenous infusion in humans, a significant insulinotropic activity was recorded [86]. Studies in β-cell lines have also suggested anti-apoptotic properties as well as β-cell proliferative effects for GIP.

Whereas GIP shares the insulinotropic properties with the other major incretin hormone GLP-1, the effects of both incretins on glucagon secretion differ largely from each other. Thus, in normoglycemic healthy volunteers, GIP even exerts a dose-dependent stimulation of glucagon release [9]. These effects are no longer detectable under conditions of hyperglycemia or in patients with diabetes. When GIP and GLP-1 were co-administered to patients with type 2 diabetes, the glucagonostatic actions of GLP-1 were fully abolished, suggesting that under normal conditions, the stimulation of glucagon release by GLP-1 overrides the glucagonostatic effect of GLP-1 [10]. In line with this, suppression of glucagon levels was found to be significantly more pronounced after intravenous compared to oral glucose administration.

**Gastrointestinal effects**

When GIP was initially discovered, a prominent role of the peptide as an inhibitor of gastric acid secretion was inferred from experiments in isolated canine stomach preparations. These studies gave rise to the initial name “gastric inhibitory polypeptide,” which was later replaced by the term “glucose-dependent
insulinotropic polypeptide” [85]. Indeed, when GIP was administered to humans with and without type 2 diabetes in vivo, no significant effect on gastric acid secretion was found [87]. Also, no effects of exogenous GIP on gastric emptying were reported. Therefore, the peptide does not seem to play a major physiologic role in the control of gastrointestinal functions in humans.

**Regulation of body weight**

A potential role for GIP as a regulator of body weight became apparent from experiments in GIP receptor knockout mice. These animals were protected from developing obesity and insulin resistance on a high-fat diet [88]. In line with these experiments, studies using various GIP receptor antagonists have demonstrated improvements in glucose homeostasis and weight gain in mice. Unfortunately, these findings have not yet been confirmed in humans or large animal species. There is also good evidence that GIP receptors are expressed abundantly on adipocytes, where the incretin is believed to induce lipoprotein lipase activity and to promote triglyceride accumulation [85]. These effects were shown to be partly mediated through increased secretion of resistin. In contrast to these experiments in rodent and cell culture models, administration of GIP to humans has had no significant effect on triglyceride or free fatty acid levels.

**Other potential actions of GIP**

In addition to the above-mentioned effects, there is some evidence for a stimulation of cortisol secretion by GIP [89]. Furthermore, GIP is believed to enhance bone formation through stimulation of osteoblast activity as well as inhibition of osteoclast resorptive activity.

**Other gastrointestinal hormones with potential incretin function**

Whereas various lines of evidence suggest that GIP and GLP-1 are the major incretin hormones in human physiology, some incretin-like activity has also been ascribed to several other gastrointestinal hormones.

**Cholecystokinin (CCK)**

Earlier studies, partly using rather nonspecific immunoassays, have reported reduced plasma concentrations of CCK in patients with type 2 diabetes, thereby suggesting a role for the peptide in the pathogenesis of diabetes. Indeed, CCK has been demonstrated to exert glucose-dependent insulinotropic effects in various experimental models [90]. However, at physiologic plasma concentrations, these effects were rather negligible, thereby questioning a true incretin function of the peptide. Furthermore, there appears to be a large interspecies variability regarding the insulinotropic effect of CCK, and experiments in humans did not reveal a strong stimulation of insulin secretion [90]. More importantly, the incretin effect was largely preserved after administration of the CCK receptor antagonist loxiglumide [91]. On that basis, even though CCK may exhibit some characteristics of an incretin hormone, it clearly cannot be considered a major physiologic mediator of the incretin effect [92].

**Secretin**

Secretin is another gut peptide with potential incretin-like properties that is secreted from gastrointestinal S cells, which are primarily located in the duodenum and upper jejunum. Secretion of secretin is stimulated by nutrient ingestion, with fat and protein acting as strong stimulators, and glucose being a rather weak stimulus. Stimulation of insulin secretion has been described in response to secretin administration in various species, but at physiologic plasma concentrations the insulinotropic effect of the peptide is rather weak, thereby questioning an incretin role of the peptide [16,92].

**Gastrin**

The hormone gastrin is secreted from G cells located in the gastric antrum and duodenum. Gastrin secretion rises modestly in response to oral glucose administration [92]. However, circulating plasma levels are also strongly affected by many other factors, such as gastric pH. Some, but not all studies have demonstrated a glucose-dependent stimulation of insulin release in response to gastrin administration. More recently, a role for gastrin in β-cell growth and regeneration has also been implicated. Overall, the magnitude of the insulinotropic effect observed with physiologic gastrin levels does not appear to be sufficient to elicit a true incretin-like activity [16,92].

**The incretin effect in patients with diabetes**

The “size” of the incretin effect, that is, the percentage contribution of incretin hormones at total insulin responses to oral glucose, has been estimated at ~35–75%, depending on the magnitude of the glucose load [6]. In a direct comparison of the incretin effect after 50 g of oral glucose between patients with type 2 diabetes and nondiabetic controls, the size of the incretin effect was 72.8 ± 6.9% in control subjects and 36.0 ± 8.8% in diabetic patients based on insulin measurements and 58.4 ± 7.6% and 7.6 ± 14.5%, respectively, based on C-peptide levels [93]. Because insulin concentrations are largely affected by hepatic insulin clearance, and oral glucose ingestion has been shown to reduce hepatic insulin clearance, it seems appropriate to primarily use C-peptide levels for the calculation of the incretin effect. On that basis it can be concluded that the incretin effect is almost completely absent in patients with type 2 diabetes. Given the importance of the incretin effect for the regulation of postprandial glucose homeostasis, it has been speculated that a loss of the incretin effects may play a central role in the pathogenesis of type 2 diabetes [94].

The incretin effect has also been examined in patients with other types of diabetes. Interestingly, when GIP and GLP-1 were
administered intravenously to patients with diabetes secondary to chronic pancreatitis, patients with MODY, LADA or early type 1 diabetes, the insulinotropic actions of the incretin hormones was reduced to a comparable extent as in patients with type 2 diabetes. In line with these observations, the size of the incretin effect was also found to be largely reduced in such patients. Taken together, these studies have lent strong support to the concept that the reduction in the incretin effect in patients with type 2 diabetes develops secondary to chronic hyperglycemia and goes along with a general decline in β-cell mass and function in these individuals [94,95].

**Potential mechanisms of incretin dysfunction in type 2 diabetes**

Various potential explanations have been expounded in order to explain the reduced incretin effect in type 2 diabetes [11,95]. One obvious factor to examine was the secretion of GIP and GLP-1. Regarding GIP, a number of studies have compared postprandial plasma concentrations between patients with and without diabetes. These studies have revealed increased, normal, or reduced GIP levels in patients with diabetes. Considering all studies together, there does not appear to be a general defect in GIP secretion [95] (Figure 11.4). If anything, a slight increase in postprandial GIP levels might occur in patients with early type 2 diabetes, but clearly the differences in GIP levels would not sufficiently explain any alterations in insulin of glucose concentrations [11].

There is some controversy regarding GLP-1 concentrations in patients with type 2 diabetes. One large cross-sectional study has reported a modest reduction in postprandial GLP-1 concentrations in patients with type 2 diabetes compared to nondiabetic control subjects. In this study, there was also a small reduction in GLP-1 levels in individuals with impaired glucose tolerance [20]. However, upon closer analysis, the defects in GLP-1 secretion in type 2 diabetes found in this study were most apparent in the late postprandial period, that is, ~120–240 min after meal ingestion. Because defects in insulin secretion in type 2 diabetes typically occur within the first 30–60 min after meal or glucose ingestion, such defect in GLP-1 levels would not plausibly explain the reductions in postprandial insulin secretion. A reduction in GLP-1 levels in patients with type 2 diabetes has also been reported in another study [21]. In this study, reduced plasma concentrations of both total and intact GLP-1 were reported. These two studies reporting lower GLP-1 levels are contrasted by a number of other studies demonstrating normal or even increased GLP-1

![Figure 11.4](image-url)
concentrations in type 2 diabetes (Figure 11.4). In a formal meta-analysis of all these trials, there were no differences in GLP-1 concentrations between patients with and without type 2 diabetes [96]. The discrepant results of these studies may be partly explained by different patient characteristics. In this regard, long diabetes duration and high glucagon plasma concentrations have been associated with modest reduction in GLP-1 levels [96]. All aspects considered, a reduction in the secretion of GIP or GLP-1 does not explain the impaired incretin effect in type 2 diabetes.

The insulinotropic effect of the two major incretin hormones has been examined in a number of experiments (Figure 11.5). For GIP, a reduced insulinotropic effect in type 2 diabetes has already been reported in initial experiments using porcine GIP preparations [86]. In a hyperglycemic clamp study directly comparing the insulinotropic effect of GIP at physiologic and supraphysiologic doses in patients with type 2 diabetes, the efficacy of GIP was reduced by ∼ 60% in the diabetic patients [97] (Figure 11.5). This finding has been subsequently confirmed in different studies. Notably, the loss of insulinotropic activity of GIP could not be overcome even by the administration of highly supraphysiologic doses [11]. The efficacy of GIP was also found to be impaired to a lesser extent in first-degree relatives of patients with type 2 diabetes, along with a general reduction in β-cell function. The mechanistic reasons underlying the loss of GIP efficacy in type 2 diabetes have not been fully elucidated [11]. Polymorphisms in the GIP receptor have recently been related to 2-h glucose concentrations during an OGTT in nondiabetic individuals. However, the effect size of this polymorphism is clearly insufficient to explain the marked abnormalities in GIP action in patients with type 2 diabetes. It has also been speculated that patients with type 2 diabetes may exhibit a reduced number of GIP receptors on pancreatic β cells, but until now, GIP receptor expression has not yet been quantified on islets from patients with type 2 diabetes. However, in rodents with diabetes a reduced expression of GIP receptors has been described, and the expression level of the GIP receptor was found to be modulated by the prevailing glucose concentrations. In line with these studies in rodent models, the insulinoergic action of the incretin hormones was found to be largely enhanced after reduction of chronic hyperglycemia by means of intensive insulin therapy in patients with type 2 diabetes. Therefore, at present the most plausible explanation for the loss of GIP efficacy in type 2 diabetes is a downregulation of GIP receptor expression secondary to chronic hyperglycemia [11].

When directly compared to nondiabetic individuals, the insulinotropic effect of GLP-1 has also been found to be reduced in patients with type 2 diabetes [97]. However, the extent of this defect is clearly less pronounced than the respective impairment

![Figure 11.5](image-url)
in GIP action. Thus, in a hyperglycemic clamp study examining individuals with and without type 2 diabetes, the insulinotropic effect of GLP-1 was reduced by ~30% in the diabetic patients [97]. Of note, this impairment in GLP-1 action appears to be less than what would be expected in such patients, assuming a ~50–65% β cell deficit in type 2 diabetes [11]. Moreover, the impairment in GLP-1 action can easily be overcome by administration of supraphysiologic doses. Thus, even though in absolute terms the insulinotropic effect of GLP-1 is reduced in type 2 diabetes, this incretin hormone still appears to stimulate insulin secretion to a much greater extent than GIP or other secretagogues. The largely preserved insulinotropic activity of GLP-1 in patients with type 2 diabetes may be secondary to the recruitment of additional β cells, as demonstrated in a series of elegant experiments in isolated β cells [48].

Although the impairment in incretin activity in type 2 diabetes primarily affects their insulinotropic effect, it is still conceivable that the actions of GIP or GLP-1 on glucagon secretion may also be affected. However, in a direct comparison between patients with and without type 2 diabetes, the glucagonostatic effect of GLP-1 was fully preserved [97]. In this regard it is worth mentioning that GLP-1 has been found to normalize glucagon levels even in patients with type 1 diabetes [51]. For GIP, a glucagonotrophic effect has been described in healthy individuals at normoglycemia [9]. These effects were no longer detectable at hyperglycemia or in patients with type 2 diabetes. Therefore, alterations in glucagon secretion do not seem to contribute to the loss of incretin activity in type 2 diabetes.

An alternative explanation to the concept of GIP receptor downregulation as a mechanism for the loss of incretin activity in type 2 diabetes would be a general decline in β-cell mass and function in these patients [11]. According to this hypothesis, the reduction in GIP activity in type 2 diabetes largely goes along with a general decline in β-cell function. Indeed, the insulinotropic effect of GIP is closely correlated to the β-cell function. Indeed, the insulinotropic effect of GIP is closely correlated to the glucose concentrations and the size of the incretin effect has been described. Thus, even though the insulinotropic effect of the incretin hormones, especially GIP, is clearly diminished in patients with type 2 diabetes, it is likely that the degree of impairment of the incretin effect is overestimated because of the experimental conditions [11].

**Summary and conclusions**

The incretin hormones GIP and GLP play a predominant role in the postprandial augmentation of insulin secretion, thereby largely contributing to the maintenance of glucose homeostasis after meal ingestion. Incretin-like properties have also been ascribed to other gut hormones, such as CCK or secretion, but the importance of these other peptides is rather negligible under physiologic conditions. In addition to their insulinotropic effects, GIP and GLP-1 exhibit numerous other actions in various other organs and tissues, such as the central nervous system, the gastrointestinal tract, and the cardiovascular system. In patients with diabetes, the incretin system is markedly altered, which appears to be primarily due to loss of insulinotropic activity of GIP. Therefore, the incretin system represents a central regulatory element in the interplay between nutrient absorption and glucose homeostasis, which is of critical importance for the physiologic control of energy homeostasis.

**References**


43 Nauck MA, Heimesaat MM, Behle K, et al.: Effects of glucagon-like peptide 1 on counterregulatory hormone responses, cognitive...
functions, and insulin secretion during hyperinsulinemic, stepped hypoglycemic clamp experiments in healthy volunteers. *Journal of Clinical Endocrinology & Metabolism* 2002;87:1239–1246.


SECTION III

Normal physiology of insulin action
CHAPTER 12  
Mechanisms of insulin signal transduction

Paul R. Langlais¹, Lawrence J. Mandarino² and W. Timothy Garvey³
¹Department of Biochemistry and Molecular Biology, College of Medicine, Mayo Clinic in Arizona, Scottsdale, AZ, USA  
²Division of Endocrinology, Mayo Clinic in Arizona, Scottsdale, AZ, USA; Center for Metabolic and Vascular Biology, Arizona State University, Tempe, AZ, USA  
³Department of Nutrition Sciences and the UAB Diabetes Research Center, University of Alabama at Birmingham, Birmingham, AL; Birmingham VA Medical Center, Birmingham, AL, USA

Key points

- The insulin signal is transduced as a web of interacting signal molecule cascades and effector systems. The final biologic action represents the net synergism of the combined facilitative, inhibitory, and complementary signaling pathways that interact with terminal functional systems.
- Insulin signaling begins with tyrosine phosphorylation of the insulin receptor and insulin receptor substrate molecules; however, signaling can be inhibited by serine phosphorylation of these same molecules mediated by inflammatory pathways.
- There are three major steps in signal transduction that facilitate divergence of activated pathways impacting multiple cellular functions and adding a high degree of regulatory flexibility: insulin receptor/receptor substrate docking molecules, PI-3 kinase, and Akt/protein kinase B.
- Insulin-mediated translocation of GLUT4 glucose transporters to the plasma membrane is a complex process involving parallel signal transduction pathways (insulin receptor/IRS/PI-3 kinase and CAP/Cbl/TC10), regulated exocytotic and retroendocytotic trafficking pathways for GLUT4-containing vesicles, and the cytoskeleton.
- Many insulin actions are mediated by effects on gene transcription due to factors that are either directly modified via insulin signaling mechanisms or indirectly via effects on glucose or lipid substrate metabolism.
- There are two major loci of defects contributing to human insulin resistance: impaired proximal signal transduction characterized by reduced tyrosine phosphorylation of insulin receptor and insulin receptor substrate molecules, and abnormalities in GLUT4 vesicle trafficking that impair translocation to the cell surface.

Introduction

Fuel homeostasis

The peptide hormone insulin is secreted from pancreatic β cells, and exerts a broad spectrum of anabolic effects in multiple tissues. In response to a meal, insulin stimulates uptake and storage of carbohydrate, fat, and amino acids, while, at the same time, antagonizing the catabolism of these fuel reserves. During periods of fasting, a fall in circulating insulin combined with increased secretion of counterregulatory hormones leads to breakdown of stored fuels and increased availability of metabolic substrates for cellular energy. In particular, insulin plays a key role in maintaining blood glucose level within a narrowly defined range despite large fluctuations in food intake. After a meal, insulin returns the organism to postabsorptive homeostasis by promoting uptake and storage of glucose as glycogen in skeletal muscle and liver, and by inhibiting hepatic glucose production via decrements in glycogenolysis and gluconeogenesis. Insulin also promotes triglyceride storage by suppressing fatty acid oxidation, increasing lipid synthesis in liver and adipose tissues, and via an antilipolytic effect in fat cells. Insulin’s actions extend to protein accretion through a diminution in protein catabolism and an increase in translation, to cell growth and differentiation, and to cell survival as a consequence of mitogenicity and anti-apoptosis.

Changing perspectives of insulin action

In mediating its pleiotropic actions, insulin engages multiple networks of signal transduction molecules and a wide array of effector systems. Insulin alters activity of enzymatic pathways, subcellular localization of enzymes, and activation state of membrane transport systems by regulating the posttranslational modification, translation, and degradation of proteins in cells and by affecting gene transcription and expression. The classical approach to the review of insulin action is to describe a linear cascade of sequentially-interacting signal transduction molecules, which then engage specific effector systems, such as membrane transport proteins or the activities of rate-limiting enzymes (e.g. glycogen synthase). This narrow perspective is myopic for three reasons. First, signal
transduction and effector functions are not clearly distinct. For example, phosphatidylinositol-3 kinase (PI-3 kinase), which is a component of linear signal transduction from insulin receptor to activation of Akt/protein kinase B (Akt/PKB) and protein kinase C (PKC) isoforms, is integrally involved in glucose transport system effector functions such as vesicle trafficking and the actin cytoskeleton [1].

Secondly, it has become clear, partly through the study of genetic disruption and hyperexpression in mice, that insulin activates numerous systems of signal transduction. These systems are at least partially redundant and are able to interact in modulating signal transduction and ultimate biologic effects. One example is the ability of Akt/PKB, activated through the PI-3 kinase “metabolic” signaling pathway, to impact mitogenic responses through regulation of gene transcription factors. Furthermore, signal transduction systems and effector systems must work in concert to mediate a given biologic effect. For example, stimulation of glucose transport requires parallel activation of PI-3 kinase and CAP/Cbl/TC10 signal pathways, as well as interaction between two effector systems, the glucose transport system (i.e., trafficking of GLUT4-containing vesicles) and effects on the cellular cytoskeleton. Thus, promulgation of insulin action is most accurately viewed as a changing pattern of network interactions involving a web of signal molecule cascades and effector systems. The final biologic action represents the net synergism of the combined facilitative, inhibitory, and complementary signaling pathways that interact with more terminal functional systems in cell biology. Certain aspects of linear insulin signal transduction are evolutionarily conserved and are quite similar in mammals and invertebrates. However, complex patterns of interactions between signal and effector systems are more pronounced in mammals and allow for greater plasticity in adaptive responses.

Thirdly, the classic descriptions of insulin signaling pathways emphasize insulin action mechanisms within certain target cells, especially fat cells. However, these actions at the level of cells and tissues affect substrate flux and release of hormones and tissue factors resulting in the coordinated function of multiple organs as whole organisms adapt to the nutritional environment. The reader is encouraged to think about insulin action in the context of systems biology and integrative physiology. This chapter will emphasize insulin action involving stimulation of glucose transport and glucose metabolism, for reasons outlined later. However, this information will complement other chapters depicting insulin’s ability to regulate protein metabolism (Chapter 16), fat metabolism (Chapter 17), and intermediary metabolism (Chapter 13).

**Insulin resistance**

Interest in defining molecular mechanisms mediating insulin action has accelerated with the recognition that insulin resistance is both a primary abnormality in the evolution of type 2 diabetes (T2DM) and is also associated with a cluster of risk factors predisposing to cardiovascular disease (the metabolic syndrome) [2]. The cardinal clinical manifestations of insulin resistance include hyperinsulinemia in conjunction with normoglycemia or hyperglycemia. Investigators have used various in vivo metabolic techniques that assess glucose uptake during insulin infusions to demonstrate that insulin resistance was due to impaired insulin action in peripheral tissues. These conclusions were subsequently confirmed with the demonstration that insulin-stimulated glucose transport was diminished in fat and muscle tissues removed from insulin-resistant subjects and studied ex vivo. Therefore, while insulin resistance could apply to any of insulin’s pleiotropic effects, the term generally applies to insulin’s ability to stimulate tissue glucose uptake and suppress hepatic glucose production since these actions are most directly relevant to its clinical manifestations (hyperinsulinemia and impaired glucose tolerance). This chapter will primarily focus on mechanisms of insulin action related to the regulation of glucose transport and metabolism. While many key aspects remain to be elucidated, this field is rapidly advancing. Over the past decade, the application of new and improved methodologies in molecular biology have facilitated the study of insulin action, resulting in new knowledge and new ways to conceptualize the effects of this powerful and interesting hormone.

**Insulin receptors**

**Ligand binding**

The insulin action sequence is initiated with specific binding to high-affinity receptors on the plasma membrane of target cells. The insulin receptor is a large transmembrane glycoprotein consisting of two α- and two β- subunits that form a heterotetramer, as depicted in Figure 12.1. The insulin receptor is synthesized from a single gene as a polypeptide preproreceptor comprised of α- and β- subunit sequences in tandem. The preproreceptor is processed into separate α- and β- subunits which then assemble as a disulfide-linked holoenzyme of stoichiometry (αβ)2 [3]. The 135 kDa α subunits, derived from the amino-terminal portion of the preproreceptor, reside entirely on the outside of the cell, tethered to the membrane via the 95 kDa β subunits which span the membrane. Insulin binds to the extracellular α subunit. Studies using chimeric receptors containing structural domains from the homologous insulin-like growth factor-1 (IGF-1) receptor have established that multiple extracellular regions including glycine-centered repeats in the α subunit establish the specificity for insulin binding [4]; however, a cysteine-rich domain has been shown to directly interact with the ligand. Insulin binding results in conformational changes that bring the α subunits closer together, and enables ATP to bind to the intracellular domain of the β subunit.

**Tyrosine kinase activity**

The β subunits possess tyrosine kinase activity [5,6]. The initial event following insulin-binding appears to be phosphorylation
Mechanisms of insulin signal transduction

Figure 12.1 Structural and functional domains of the insulin receptor. Schematic diagram of the insulin receptor showing structural landmarks (left side) and functional domains (right side). (Numbering of amino acids based on translation of the alternatively spliced exon 11.) (For a color version of this figure, please see color plate section.)

of one β subunit catalyzed by a specific domain in the other β subunit within the same heterotetramer, a reaction that has been termed “transphosphorylation” or autophosphorylation [7]. Autophosphorylation of the kinase renders it considerably more active, allowing the accelerated transfer of phosphate to other tyrosine sites within the same receptor as well as exogenous substrates [8]. Ligand-dependent stimulation of the β-subunit tyrosine kinase activity is critical for promulgation of the insulin signal. There is an ATP-binding domain in the β subunit, and at least six tyrosine residues undergoing phosphorylation that have been shown in mutagenesis experiments to serve different roles in insulin signaling, as shown in Figure 12.1. These phosphotyrosines lie within three functional groups. Tyrosine phosphorylation sites at positions Y1158, Y1162, and Y1163 are essential for mediating an increase in subunit tyrosine kinase activity and signal transduction, and lie in the active loop of the kinase catalytic domain. When phosphorylated, a juxtamembrane tyrosine at position Y972 becomes part of a recognition motif and provides a docking site for several intracellular substrates. Y972 is required to assure sufficient stability of the receptor/substrate complex for substrate phosphorylation. Y972 also lies within a domain that mediates endocytosis of ligand-bound insulin receptors. Phosphorylation sites Y1328 and Y1334 near the COOH terminus are not essential for stimulation of glucose transport but may affect the sensitivity of Ras/MAP kinase pathway activation, and, so, are involved in the receptor’s mitogenic responses [9]. Phosphorylation of serine and threonine residues within the insulin receptor also occurs following insulin binding, and, as described later, phosphorylation of these residues by intracellular serine kinases can diminish receptor tyrosine kinase activity; it represents a regulatory feedback mechanism [10] and can induce insulin resistance under pathophysiologic conditions.

Insulin-like growth factors

The insulin receptor is similar in structure to IGF-1 and other growth factor and cytokine receptors (all (αβ)2 stoichiometry), and IGF-II receptors ((αβ) stoichiometry), in having an extracellular ligand-binding domain that activates an intracellular tyrosine kinase domain [2,3,11]. The respective ligands, insulin and IGF-1, are also closely related hormones from an evolutionary perspective, but have widely different functions in regulating metabolism and growth, respectively. Nevertheless, it has been recognized for some time that these hormones can elicit similar effects to one another, if they are present at high concentrations. This suggested that they could bind to one another’s receptors, with differing affinities. In fact, both IGF-1 and IGF-II can bind to cell surface insulin receptors. The affinity of IGF-1 for the insulin receptor is 100- to 1000-fold less than that for insulin. Nevertheless, circulating IGF-1 concentrations can be 100-fold higher than insulin, such that there is some potential for IGF-I binding and action through the insulin receptor. IGF-II binds insulin receptors during embryogenesis in rodents, and promotes fetal growth. This capability may be associated with expression of an alternatively spliced form of the insulin receptor with high affinity for IGF-II [12]. Alternative
splicing in exon 11 determines the insertion or deletion of 12 amino acids near the COOH terminus of the α subunit. Isoform A lacking the 12 amino acids has high affinity for IGF-II, predominates during fetal development, and promotes growth as a consequence of IGF-II binding. Isoform B containing the 12 amino acids predominates postnatally and is activated mainly by insulin. Some evidence supports the contention that dysregulated expression towards the fetal pattern could occur in adult tissues and result in insulin resistance [12,13].

**Hybrid receptors**

Because of the heterotetrameric nature of the insulin and IGF-1 receptors, it was natural to ask whether individual dimers could be formed from “hybrid” receptors, consisting of one α/β insulin receptor dimer and one α/β IGF-1 receptor dimer. Such hybrid receptors were indeed demonstrated to exist [14–16], and it appears likely that these hybrid receptors form stochastically, such that if half of the dimers present were insulin receptor dimers, and half were IGF-1 receptor dimers, half of the mature receptors present in the membrane would be hybrid insulin/IGF-1 receptors, with about one quarter each “pure” insulin or IGF-1 heterotetramers. This raises the question of what the main ligand is for such hybrid receptors, and what downstream effects would be triggered when the hybrid receptors are occupied. Most evidence would suggest that hybrid receptors bind IGF-1 with greater affinity than insulin and mediate IGF-1-like effects [17].

The implication of these findings is that the ratio of IGF-1 to insulin receptor dimer subunits could have an important influence on the number of insulin receptor heterotetramers that exist in the plasma membrane and can actually function to mediate insulin’s metabolic effects. Increased IGF-1 receptor expression can produce insulin resistance via increasing the proportion of hybrid receptors, essentially hijacking insulin receptor dimers to serve a function in IGF-1 signaling [16]. For example, aldosterone can induce vascular insulin resistance by this mechanism [18], and IGF-1 receptor expression can modulate insulin sensitivity and nitric oxide availability in endothelium [19]. Hybrid insulin/IGF-1 receptors may also play a role as growth factor receptors in myeloma cells [20] and endometrial cancer cells [21], as well as affecting the function of platelets [22]. The role of hybrid insulin/IGF-1 receptors in regulating insulin action and disease processes is an understudied area.

**Insulin receptor substrate molecules**

**Receptor substrate/docking proteins**

Following insulin binding and receptor autophosphorylation, the next committed step in signal transduction is tyrosine phosphorylation of intracellular proteins. To accomplish this, autophosphorylation of the β subunit mediates noncovalent but stable interactions between the receptor and intracellular substrate proteins, and this positions these molecules for tyrosine phosphorylation by the activated insulin receptor kinase [26–28]. Several proteins are rapidly phosphorylated on tyrosine residues by ligand-bound insulin receptors, including six insulin receptor substrate proteins (IRS-1, IRS-2, IRS-3, IRS-4, IRS-5, IRS-6) [29–33], Src-homology collagen proteins (SHC) [34] and, growth factor receptor bound-2 (Grb-2) associated binder-1 (Gab-1) [35], signal-regulatory protein (SIRP) family members [36,37], the CAP/c-Cbl complex [38,39], an adapter protein with a PH and SH2 domain (APS) [40], signal transducer and activator of transcription 5B (STAT5B) [41], and proteins referred to as downstream of kinase (DOK) 1-6 [33,42].

There is no known enzymatic activity associated with insulin receptor substrate molecules; however, their primary structure is noteworthy for multiple sites capable of interaction with other proteins. The receptor substrates are characterized by a representative architecture, particularly with respect to three functional domains. These include an N-terminal pleckstrin homology (PH) domain, an intermediate phosphotyrosine binding (PTB) domain, and a carboxy-terminal region of variable length that contains the multiple tyrosine and serine phosphorylation sites [26–28]. The PH domain helps position the insulin receptor for coupling with IRS, possibly
by binding to charged headgroups of certain phosphatidylinositides in adjacent membrane structures. PTB domains recognize the phosphotyrosine in an amino acid sequence asparagine-proline-any amino acid-phosphotyrosine (NPXpY), and this motif is present in the insulin receptor and other receptors with tyrosine kinase activity such as the IGF-1 receptor. In this manner, the PTB domain enables receptor substrates to directly bind Y972 of the insulin receptor β subunit. All of the receptor substrate molecules have PH domains except SHC, and all have a PTB domain except Gab-1. These domains facilitate binding of the receptor substrates to the phosphorylated insulin receptor, allowing for phosphorylation of tyrosine residues in their COOH-terminal domains.

In the carboxy end of insulin receptor substrate molecules, the phosphotyrosines provide binding sites (pYMXM) for other proteins containing Src homology-2 (SH2) domains. This term relates to homology with pp60^src, the src oncogene product and first protein kinase demonstrated to possess activity towards tyrosine residues [43]. Noncovalent attachment of SH2-containing signaling molecules allows receptor substrates to function as docking proteins that appose proteins in the transduction cascade. Receptor substrates thus propagate the insulin signal via the docking, apposition, interaction, and activation of downstream signal molecules, rather than as a consequence of any intrinsic enzymatic activity. This system provides for extensive plasticity and regulation. For example, the multiple receptor substrate molecules exhibit differences in tissue-specific expression and in their capacity to activate various downstream signal pathways. In addition, the receptor substrates can be acted upon by multiple receptor tyrosine kinases, such as IGF-1, interleukin, and integrin receptors, as well as serine kinases (see later), in a manner that allows for modulation of insulin signal and integration of responses to extracellular factors. Thus, these docking molecules provide for divergence, heterogeneity, and regulation of insulin signal transduction processes.

**Insulin receptor substrate (IRS) family of docking proteins**

IRS-1 and IRS-2 exhibit a wide range of tissue expression, including muscle, fat, liver, and pancreatic islets, although their relative levels can vary in different tissues [27–30]. IRS-3 is expressed in adipose tissue, fibroblasts, and liver cells, while IRS-4 has been detected in brain, thymus, and embryonic kidney [31,32]. IRS-5 is ubiquitously expressed but most abundant in kidney and liver. IRS-6 expression is highest in skeletal muscle [33]. Knockout mouse models have highlighted functional differences among the IRS protein family. Genetic ablation of IRS-1 leads to severe growth retardation reflecting a decrease in growth-promoting effects of insulin and other factors which signal through IRS-1 [44,45]. These mice were also insulin resistant but did not develop overt diabetes. On the other hand, ablation of IRS-2 produced insulin resistance, impaired insulin secretion, and overt diabetes in mice that were normal in size [46]. The IRS-2 knockout mice were observed to have decreased β-cell mass and insulin content, indicative of a trophic role for insulin signaling through IRS-2 in β-cell development. Genetic ablation of IRS-2 recapitulates the pathogenesis of T2DM in that diabetes arises only when insulin resistance and insulin secretion are both impaired in these knockout mouse models. IRS-3 knockout mice do not have an obvious phenotype, and IRS-3 has not been detected in the human genome. IRS-4 null mice also appear normal with the exception of reduced fertility. While IRS-5 and IRS-6 undergo insulin-stimulated tyrosine phosphorylation, their relevance to insulin action has been questioned, as both isoforms show very weak affinity for the insulin receptor [47]. Therefore, IRS-1 and IRS-2 are the IRS isoforms most critically important in glucose homeostasis. The consensus that has arisen from these studies is that, while the IRS family members to some extent represent duplicative pathways for insulin signal transduction, IRS-1 functions as the principal IRS in skeletal muscle, and that IRS-2 predominates in liver and β cells where insulin action is required for normal β-cell growth and development.

In addition to PH and PTB domains assuring close apposition to the insulin receptor, IRS-2 (but not other IRSs) contains another region that interacts with the phosphorylated regulatory loop of the insulin receptor kinase, and this region has been designated the kinase regulatory loop binding domain (KRLB) [41,48,49]. Crystal structure analysis of a portion of the KRLB domain indicates that this region acts to restrict tyrosine phosphorylation of IRS-2 [50]. While the KRLB domain could confer some measure of signal specificity, it is unclear whether the presence of the KRLB explains the differences in phenotype between IRS-1 versus IRS-2 knockout mice.

IRS proteins are a key locus for regulation of insulin action. One mechanism occurs at the level of IRS protein expression; for example, IRS-1 and IRS-2 proteins are decreased by hyperinsulinemia [51]. Cell loss of IRS could occur through accelerated protein degradation due to induction of ubiquitin-mediated degradation of IRS-1 and IRS-2 by SOCS proteins [52], or by inhibition of IRS gene transcription. Regardless of the mechanism, decreased levels of IRS proteins in hyperinsulinemic states, coupled with downregulation of the insulin receptor itself, can contribute to the insulin resistance in diabetes [53]. The function of IRS proteins can also be negatively regulated by serine/threonine kinases and protein tyrosine phosphatases such as PTP1B and SHP2 [54] as described later. In addition, IRS-1 can be posttranslationally modified by either O-linked N-acetylglucosamine adducts (O-GlcNAc) on serine/threonine residues under hyperglycemic conditions [55], or by S-nitrosylation as a consequence of nitric oxide generation [56]. These modifications induce the proteasomal degradation of IRS-1 and insulin resistance [57].

**Other receptor substrates**

Like IRSs, SHC proteins are phosphorylated in response to insulin, lack any known catalytic activity, and interact with
SH-2 domain-containing proteins through their tyrosine phosphorylation sites [34]. A phosphorylation site on SHC binds Grb-2, which can then lead to the activation of the Ras/MAP kinase mitogenic signaling pathway. Therefore, SHC constitutes an additional substrate for the insulin receptor kinase and transduces the insulin signal via its protein docking properties. There are at least three known SHC isoforms (46, 52, and 66 kDa), and all contain an amino terminal phosphotyrosine-binding domain, a central region that is homologous to the α1 chain of collagen, and a carboxy terminal SH-2 domain.

Gab-1 contains only a PH domain at the amino terminus and a domain containing several tyrosine phosphorylation sites, but lacks a PTB domain [35]. In this case, the PH is sufficient for positioning Gab-1 adjacent to the insulin receptor and enabling phosphorylation, although a tight anchor is not possible due to the lack of a PTB domain. Gab-1 is most heavily phosphorylated by the ligand-activated epidermal growth factor receptors, another receptor containing tyrosine kinase activity, and less well by activated insulin receptors.

The DOK family consists of DOK1 through 7 and are similar to IRS family members in domain architecture [33,42,58]. DOK proteins appear to have functional effects in lymphocytes, myeloid cells, muscle cells, and neurons, and can be phosphorylated by both membrane-associated and cytoplasmic tyrosine kinases. DOKs are docking proteins that have been observed to associate with Ras GTPase-activating protein (RasGAP), Nck, c-Abl, and the insulin receptor, and in cultured cells can be involved in cytoskeletal reorganization. DOK1 (p62dok) has been shown to negatively regulate insulin-stimulated activation of AKT [59], although the role of the remainder of the DOK family proteins in mediating the biologic effects of insulin is not well understood.

Proteins that dock with receptor substrate molecules

SH2-containing proteins, which bind to phosphotyrosine motifs and dock with insulin receptor substrate proteins, include proteins with enzymatic activity such as PI-3 kinase, Fyn (tyrosine kinase), Csk (tyrosine kinase), and SHP-2 (phosphotyrosine phosphatase), and other adapter proteins such as Grb-2, Crk, APS, and Nck [27,28]. In addition to SH2 domains, the adapter proteins also contain SH3 domains, which bind to proline-rich sequences in other proteins with consensus sequence PXXP. In this way, the adapter proteins associate with receptor substrates via their SH2 domains and bring with them other proteins bound to their SH3 domains. The SH3-bound proteins represent downstream signaling molecules and catalytic subunits, which participate in transduction of the insulin signal. The proteins that dock with receptor substrates, whether binding directly or via adapters, have different enzymatic activities that activate specific downstream molecules as a result of their juxtaposition on receptor substrate docking molecules. In this way, insulin receptor substrates provide the first major point of divergence of insulin signal transduction pathways, leading to activation of the mitogenic (Ras/MAP kinase), and metabolic (PI-3 kinase) pathways. This is illustrated in Figure 12.2.

Ras/MAP kinase pathway: mitogenic signaling

One component of signal divergence emanating from IRS docking proteins is the engagement of the Ras/MAP kinase mitogenic signaling pathway (Figure 12.2). One of the SH-2 domain-containing proteins that docks with IRS is Grb-2, a small cytosolic adapter protein. Grb-2 also contains an SH-3 binding domain which binds proteins via interaction with proline-rich sequences, and one of these proteins is SOS (mamalian homologue of the Drosophila son-of-sevenless protein), a GDP/GTP exchange factor. Following insulin stimulation, Grb-2 is able to bind IRS via its SH-2 domain, and position SOS for activation of the Ras signaling pathway. SOS facilitates GTP activation of membrane-bound Ras, the 21 kDa small molecular weight GTPase, which has been demonstrated to play a major role in cell growth and oncogenesis. The GTP-bound form of Ras complexes with and activates Raf-1 kinase, and initiates a cascade leading to sequential phosphorylation and activation of MAP kinase kinase, MAP kinase, and p90Rsk. Insulin receptors can also mediate activation of the Ras/MAP kinase pathway through another substrate docking molecule, SHC [60]. Similar to IRSs, SHC activates Ras and the MAP kinase pathway by forming a complex with Grb-2/SOS in response to insulin. Whether activated through IRS or SHC, MAP kinase translocates into the nucleus where it phosphorylates transcription factors mediating the mitogenic and growth promoting effects of insulin [61]. Phosphorylation of nuclear transcription factors modulates their DNA binding properties and ability to regulate gene transcription. For example, p90Rsk phosphorylates c-fos, and MAP kinase phosphorylates Elk-1, increasing transcriptional factor activity in both instances. The MAP kinase cascade is also one of the pathways with the potential to stimulate glycogen synthase since p90S6 kinase is able to activate the glycogen-associated protein phosphatase-1 (PPG-1) which in turn dephosphorylates and activates glycogen synthase. In this way, the MAP kinase pathway has the potential to interact with metabolic signaling pathways (see later). However, the MAP kinase pathway is not necessary for stimulation of glucose transport, and is not viewed as being critically related to the metabolic effects of insulin. For example, inhibition of the pathway in adipocytes using dominant negative forms of Ras [62] or inhibitors of MAPKK [63] will block transcriptional effects associated with the MAP kinase pathway, but will not interfere with insulin stimulation of either glucose transport or glycogen synthesis.
PI-3 kinase pathways: regulation of metabolism and gene expression

Phosphatidylinositide-3 kinase

Signal transduction for insulin’s metabolic effects also diverges from insulin receptor substrate proteins and proceeds via the PI-3 kinase pathway (Figure 12.2). The first committed step involves type 1A PI-3 kinase, a heterodimer consisting of a p85 regulatory subunit and a p110 catalytic subunit [64]. In quiescent cells, the regulatory subunit maintains a state of low activity for the catalytic subunit. Upon insulin-mediated tyrosine phosphorylation, IRS-1 and IRS-2 (and other docking proteins) bind PI-3 kinase via interaction with the SH-2 domain on the p85 regulatory subunit [65]. Insulin stimulation increases the amount of PI-3 kinase associated with IRS, and the binding process increases the specific activity of the p110 catalytic subunit. Activation of PI-3 kinase is critical for transducing the metabolic actions of insulin, including stimulation of GLUT4 translocation to the plasma membrane with subsequent stimulation of glucose transport activity. This has been demonstrated in several lines of investigation. Treatment of adipocytes with (i) inhibitors of PI-3 kinase including wortmannin or LY-294002, (ii) hyperexpression of dominant negative mutants of PI-3 kinase that interfere with endogenous wild-type enzyme function, and (iii) microinjection of neutralizing antibodies to the p110 catalytic subunits all result in abrogation of insulin’s biologic effect [9]. Conversely, hyperexpression of wild-type PI-3 kinase or a constitutively active mutant leads to stimulation of both glucose transport activity and GLUT4 translocation from 50–100% of insulin’s maximal effect depending on the study [66].

Several proteins can serve as the regulatory subunit of PI-3 kinase [28]. Multiple isoforms arise through alternative splicing of the p85α gene (p85α, p55α/AS53, and p50α each occurring with and without a spliced insert), and two other isoforms represent products of different genes (p85β, p55β). All regulatory subunit isoforms associate with IRS phosphotyrosines and mediate activation of the associated p110

Figure 12.2 Metabolic and mitogenic signaling pathways in insulin action. Abbreviations: DAG, diacylglycerol; IP3, inositol-3-phosphates; PI(3,4)P2, phosphatidylinositide 3,4 diphosphate; PI(3,4,5)P3, phosphatidylinositide 3,4,5 triphosphate; PI-3 kinase, phosphatidylinositide-3 kinase; PDK1, phosphoinositide-dependent kinase 1; GSK3, glycogen synthase kinase 3; GS, glycogen synthase; GLUT-4, glucose transport protein 4; IRS, family of insulin receptor substrate proteins; SHIP2, SH2-containing inositol phosphatase 2; Grb2, growth factor receptor bound 2; SOS, mammalian homologue of Drosophila son-of-sevenless protein; Akt/PKB, protein kinase B; Shc, adapter protein with homology with Src and collagen; Ras, rat sarcoma protein; Gap, GTPase activating protein; MAP kinase kinase, mitogen activated protein kinase kinase; MAP kinase, mitogen activating protein kinase; PP1G, glycogen-associated protein phosphatase 1. (For a color version of this figure, please see color plate section.)
catalytic subunit. While p85α is ubiquitously expressed and is the dominant form in many tissues, the other isoforms exhibit differential tissue-specific expression. The various isoforms of the regulatory subunit could serve to modulate or direct signal transduction via different affinities for IRS proteins or differences in subcellular compartmentalization. In fact, studies have further illuminated novel mechanisms of regulation that involve different regulatory subunits of PI 3-kinase, and it has also become clear that the relative abundance of catalytic and regulatory subunits can influence insulin action [67–73]. For example, investigators have explored the roles of p85α (Pik3r1) and p85β (Pik3r2) regulatory subunits by using brown adipose cell lines with disruption of these genes [72]. These studies revealed key differences in the way the p85 regulatory subunit isoforms influence downstream signaling effects. Other studies have demonstrated that the balance between p85α regulatory and p110 catalytic subunit abundance regulates both insulin and IGF-1 signaling [71]. These data suggest that lowering the abundance of the p85α regulatory subunit can enhance insulin sensitivity and could present a new therapeutic target for insulin resistance.

The function of the activated PI-3 kinase is to phosphorylate the inositol ring in plasma membrane glycolipids at the D-3 position, thereby converting phosphatidylinositol 4,5 bisphosphate to phosphatidylinositol 3,4,5 trisphosphate (PI(3,4,5)P3), and to a lesser extent phosphatidylinositol 4-phosphate to phosphatidylinositol 3,4 biphosphate (PI(3,4)P₂) [74,75]. In skeletal muscle, the PI-3 kinase inhibitor, wortmannin, has been shown to prevent both the formation of 3′ phospholipids and block stimulation of glucose uptake in response to insulin. While the proximal physiologic targets for the D-3 phosphoinositides have not been fully elucidated, phosphorylation of the inositol ring at the 3′ position in membrane glycolipids recruits certain signaling proteins with pleckstrin homology domains to the plasma membrane. Binding to membrane-associated phosphoinositides both activates these proteins and positions them for downstream signal transduction. In addition, phospholipase C may participate in the release of the 3′ inositol phosphate moieties and may be activated by insulin through an IRS-independent mechanism [76,50], although siRNA-based protein knockdown of phospholipase Cγ did not affect insulin signaling to the glucose transport system in 3T3-L1 adipocytes [77].

While the role of PI-3 kinase in stimulation of glucose transport has been widely recognized, it has also become clear that PI-3 kinase activation is necessary for mediating multiple other metabolic effects of insulin including regulation of gene transcription. Thus, PI-3 kinase is the second major step, downstream of insulin receptor substrates, providing for broad divergence of insulin signal transduction, as illustrated in Figure 12.3.

This is well demonstrated by experiments wherein suppression

---

**Figure 12.3** Three major steps providing for functional divergence of insulin signal transduction. The insulin receptor substrate docking proteins, PI-3 kinase, and Akt/PKB are sequentially activated by insulin, but then each factor engages multiple downstream signal molecules and pathways resulting in divergence, capacity for interaction, and complementarity of insulin signaling pathways. Abbreviations: APS, adaptor protein containing PH and SH2 domains; IRS, insulin receptor substrate protein; Gab-1, GRB-2-associated binding protein 1; Shc, adapter protein with homology with Src and collagen; PI(3,4,5)P₃, phosphatidylinositol 3,4,5 triphosphate; mTOR, mammalian target of rapamycin; p70S6K, p70 S6 kinase; Akt/PKB, protein kinase B; aPKC, atypical protein kinase C isoforms; GSK3, glycogen synthase kinase-3; BAD, Bcl2 antagonist of cell death; NO, nitric oxide. (For a color version of this figure, please see color plate section.)
of PI-3 kinase has been shown to impact a host of insulin bioeffects including stimulation of glucose transport, glycogen synthesis, and glycolysis; promotion of lipogenesis via effects on fatty acid synthase and acetyl CoA carboxylase; suppression of gluconeogenesis; stimulation of protein synthesis; stimulation of DNA synthesis; attenuation (e.g. phosphoenolpyruvate carboxykinase (PEPCK) and glucose-6-phosphatase) and induction (e.g. glucose-6-phosphate dehydrogenase and hexokinase II) of gene transcription; cell survival; and cell cycle progression. In order to propagate many of these effects, PI-3 kinase activates three serine-threonine kinases and their downstream pathways, namely, Akt or protein kinase B (Akt/PKB), atypical PKC isoforms (ζ,λ), and phosphoinositide-dependent kinases (PDK1 and PDK2).

3-Phosphoinositide-dependent protein kinase-1 (PDK1)
Insulin-mediated activation of PI-3 kinase and generation of PI(3,4,5)P3 in membrane glycolipids results in the recruitment of AGC superfamily members (protein kinase A/protein kinase G/protein kinase C family) which contain pleckstrin homology domains. This activates these serine/threonine protein kinases and positions them for downstream signal transduction. With regard to insulin signaling, the most critical of the AGC kinases activated by interaction with PI(3,4,5)P3 is the PDK1, which participates in the downstream activation of Akt/PKB and aPKCs.

The AGC superfamily includes Akt/PKB and atypical PKCs.

Akt/protein kinase B
There are three isoforms of Akt/PKB; Akt1 is expressed in a variety of tissues, while Akt2 is most highly expressed in adipocytes and Akt3 in certain cultured cell models [78–81]. Phosphorylation and activation of Akt/PKB isoforms involve both 3′ phosphoinositides and the action of PDK. 3′ Phosphoinositides in membrane glycolipids bind directly to the PH domain of Akt/PKB, and this process results in the uncovering of phosphorylation sites. At the same time, 3′ phosphoinositides activate PDK1, which in turn phosphorylates the activation loops of Akt/PKB on Thr308, enhancing the activity of the kinase [82]. However, Akt/PKB requires a second phosphorylation on Ser473 for full activation, and this could be accomplished in either of three ways. This could arise through the action of another putative phosphoinositide-dependent protein kinase, PDK2, or could be accomplished by PDK1 following a change in substrate recognition that results from threonine 308 phosphorylation, or may be the consequence of autophosphorylation. Recently, it has been suggested that rapamycin insensitive companion of mTOR (rictor) is necessary for the Ser473 phosphorylation step [83].

Regardless of the mechanism of phosphorylation, there are several lines of evidence that implicate a role for Akt/PKB activation in the stimulation of glucose transport. Expression of constitutively active membrane-bound forms of Akt results in persistent translocation of GLUT4 to the plasma membrane in muscle, fat, and cultured cell systems [84]. Interestingly, insulin stimulation is accompanied by association of Akt2 with GLUT4-containing vesicles in rat adipocytes resulting in phosphorylation of constituent proteins [85]. Conversely, cellular introduction of neutralizing antibodies, substrate peptides, and dominant negative mutant blocked insulin-stimulated GLUT4 translocation in adipocytes, albeit by only ~50%. These experiments constitute strong evidence for a role of Akt/PKB in insulin stimulation of glucose transport. However, these results are somewhat controversial because other investigators demonstrated that a dominant-interfering mutant, while blocking insulin-stimulated protein synthesis, did not inhibit GLUT4 translocation [86]. Furthermore, in muscle from insulin-resistant diabetic humans, insulin-stimulated Akt phosphorylation was intact despite gross impairment in glucose uptake [87]. These discrepant results regarding a role for Akt/PKB do not exclude a role in the stimulation of glucose transport, perhaps working in concert with parallel activating pathways (e.g. PKCζ, CAP/Cbl/TC10). Mice deficient in Akt2 exhibit mild insulin resistance in muscle, glucose intolerance, and an inability to suppress hepatic glucose production, whereas Akt1 gene disruption produces growth reduction but has little effect on glucose tolerance [88,89]. These results are indicative of an important role for Akt/PKB in insulin regulation of glucose homeostasis and also suggest that Akt1 and Akt2 may compensate for each other’s absence to varying degrees. Nevertheless, the downstream targets of Akt/PKB connecting this signaling pathway to stimulation of the glucose transport effector system have not been identified with certainty.

Akt/PKB activation is involved in multiple other insulin responses [89]. One of the first substrates identified for Akt/PKB was a serine 9 residue in glycogen synthase kinase 3β (GSK3β) [90,91]. Thus, one pathway that could contribute to insulin-stimulated glycogen synthesis involves Akt/PKB-mediated phosphorylation and inactivation of glycogen synthase kinase 3β resulting in activation of glycogen synthase activity (see Chapters 13 and 14 concerning regulation of carbohydrate metabolism). Akt/PKB is also capable of phosphorylating and activating mammalian target of rapamycin (mTOR), which in turn phosphorylates 4E-BP1/PHAS-1, leading to release of translation initiation factor eIF-4E, and recruitment of mRNA to the 40S ribosomal subunit [92]. The ribosomal S6 protein (p70S6K) is phosphorylated and activated either through PDK1, Akt/PKB, or mTOR, and similarly enhances translation of select populations of mRNA by increasing their interaction with the 40S ribosome [93]. In this way, Akt/PKB helps mediate insulin’s effect to stimulate protein synthesis (see Chapter 16: Insulin regulation of protein metabolism). Another apparent action is the phosphorylation and activation of endothelial NO synthase leading to increased NO generation and vasodilatation [94]. Additional substrates for Akt/PKB include BAD and caspase-9, which are involved in anti-apoptotic effects and alteration of cell survival pathways [95]. Finally, Akt/PKB phosphorylates forkhead transcription...
factors [96,97] altering expression patterns of genes involved in carbohydrate and lipid metabolism. In fact, the gene regulatory effects of MAP kinase and Akt/PKB-activated forkhead transcription factors generally oppose each other. In this way, Akt/PKB represents an important mechanism for “cross-talk” between the PI-3 kinase pathway, and other pathways regulating gene transcription and mitogenic effects, in the web of insulin signal transduction (Figure 12.3).

**Protein kinase C**

The IRS/PI-3 kinase pathway also activates PKCζ and PKCα, two serine-threonine protein kinase C isoforms in the “atypical” class that is not activated by calcium binding, diacylglycerol, or phorbol esters [98,99]. PI-3 kinase-dependent activation of PKCζ/λ has been observed in multiple target tissues and cultured cell models in response to insulin [100]. The activation of PKCζ/λ occurs proximal to Akt/PKB as a result of direct interaction with 3′ phosphoinositides and/or through phosphorylation and activation by PDK1, which phosphorylates PKCζ on Thr410 enhancing the activity of the kinase [82]. This is relevant to the stimulation of glucose transport since overexpression of constitutively active forms of PKCζ or PKCα increase glucose transport activity and GLUT4 translocation by 50–100% of the extent observed in response to maximal insulin [100]. Also, expression of dominant-interfering PKC mutants, lacking a critical lysine in the kinase domain or at the site phosphorylated by PDK1, inhibits insulin's ability to promote glucose transport and GLUT4 translocation by ~50%. As is common in the insulin action field, discrepant results have also been reported where overexpression of wild-type or dominant negative mutant forms in adipocytes did not affect basal or insulin-stimulated glucose transport [101]. This type of recurring controversy indicates that hyperexpression and gene knockout experiments should be interpreted cautiously especially when expression levels are altered well above or below the physiologic range, and also reflects overlapping and complementary pathways that comprise the web of insulin signal transduction. Nevertheless, on balance, available data indicate that PKCζ/λ does participate in insulin’s metabolic signaling pathways [100]. As is the case for Akt/PKB, the downstream targets of PKCζ/λ have not been fully identified, and this will be necessary for understanding of the relative participation of these two signaling proteins in the stimulation of glucose transport and other metabolic effects. Atypical PKCs can also mediate insulin simulation of general protein synthesis in a rapamycin-insensitive manner [102].

**Degradation of the PI-3 kinase signal**

At least two different types of phosphatases can degrade phosphoinositides leading to deactivation of the PI-3 kinase signal [75]. Overactivity of these enzymes has the potential for inhibiting insulin signal transduction distal to the generation of 3′ phosphoinositides and causing insulin resistance. Src-homology 2-containing inositol phosphatases (SHIP1 and SHIP2) dephosphorylate the 5′ position of PI(3,4,5)P3 to form PI(3,4)P2. Genetic disruption and loss of SHIP2 results in a marked increase in insulin sensitivity, implying that PI-3 kinase-mediated formation of PI(3,4,5)P3, rather than PI(3,4)P2, is the critical 3′ phosphoinositide in insulin signaling. Another phosphatase, phosphatase and tensin homologue (PTEN), dephosphorylates the 3′ position converting PI(3,4,5)P3 to PI(4,5)P2. Some human cancers are associated with loss of PTEN expression suggesting that unrestrained production of 3′ phosphoinositides could result in oncogenesis.

**PI-3 kinase-independent pathways for stimulation of glucose transport**

**The CAP/Cbl/TC10 pathway**

Substantial evidence confirms the presence of PI-3 kinase independent pathways for stimulation of glucose transport. It had long been clear that other growth factor receptors (i.e., the platelet-derived growth factor (PDGF) receptor, cytokine receptors such as IL-4, and certain integrins) activate PI-3 kinase to the same extent as the insulin receptor with generation of PI(3,4,5)P3, yet still do not stimulate glucose transport [103,104]. Since stimulation of transport was blocked by PI-3 kinase inhibitors (e.g. wortmannin), investigators concluded that PI-3 kinase is necessary but not sufficient for this key metabolic effect of insulin [105]. The additional factor that endows the insulin receptor with the specific capacity to stimulate glucose transport could include another parallel signaling pathway, compartmentalization of the PI-3 kinase response, or the utilization of specific signal protein isoforms in insulin target cells.

Multiple observations support a second parallel complementary pathway for insulin-stimulated glucose transport [105,106]. For example, adenovirus-mediated hyperexpression of dominantly interfering PH and PTB domains blocks insulin receptor/IRS interaction and inhibits some responses involving PI-3 kinase, such as membrane ruffling and DNA synthesis, without any effect on GLUT4 translocation. Similarly, blockade of PI-3 kinase activation, via prolonged pretreatment with PDGF or overexpression of GSK3β, inhibits association of PI-3 kinase with IRS proteins without affecting stimulation of glucose transport or GLUT4 translocation. In addition, specific mutations in the insulin receptor have been identified that exhibit a near normal ability to activate PI-3 kinase but completely fail to induce GLUT4 translocation [107]. Finally, introduction of a cell permeable analogue of PI(3,4,5)P3 does not have any effect to stimulate basal glucose transport, but in the presence of wortmannin plus insulin does elicit a maximal glucose transport response [108]. These data indicate that insulin signaling requires a separate wortmannin-insensitive pathway, which must operate in parallel with the PI-3 kinase pathway, for stimulation of glucose transport.
Evidence has accumulated strongly supporting the CAP/Cbl/TC10 pathway as the complementary pathway for insulin-stimulated glucose transport [105,106], as illustrated in Figure 12.4. This pathway was elucidated as a linear sequence of interacting factors using technologies such as the yeast two-hybrid system and co-immunoprecipitation of interacting factors. The pathway diverges from the PI3-kinase pathway at the level of the insulin receptor kinase, which mediates tyrosine phosphorylation of Cbl proto-oncogene through a process that does not involve IRSs. c-Cbl phosphorylation does require another adapter protein, APS, which binds the insulin receptor via SH2 domain-phosphotyrosine interaction and couples the receptor to Cbl [109]. c-Cbl is also complexed with another adapter protein, the c-Cbl associating protein (CAP), which helps facilitate recruitment of CAP/Cbl to the microdomain of the insulin receptor [110]. Tyrosine phosphorylation of Cbl results in disengagement of the CAP/Cbl complex from the insulin receptor.

CAP contains three tandem SH3 domains in the COOH-terminus that binds to a proline-rich domain of Cbl, and, in the NH2-terminal region, also contains a sorbin homology (SoHo) domain. After disengagement of CAP/Cbl from the insulin receptor, the CAP SoHo domain binds to flotillin in caveolin-containing lipid raft domains of the plasma membrane [111]. Lipid rafts are plasma membrane domains, enriched in cholesterol, glycolipids, and sphingolipids, which coordinate signaling events by accumulating specific protein constituents. One class of lipid rafts contains caveolin proteins in invaginated structures referred to as caveolae ("little caves"). Flotillin is anchored in lipid raft domains via its interaction with caveolin proteins. Thus, CAP is responsible for relocation of the CAP/Cbl complex to lipid raft domains. Once bound to flotillin, tyrosine phosphorylated Cbl presents a recognition site for recruitment of the CrkII-C3G complex to the lipid raft. CrkII is a small adapter protein that contains an SH2 domain that binds phosphorylated Cbl, and SH3 domains that interact with a proline-rich region of C3G. C3G is a guanyl-nucleotide releasing protein (C3G) for TC10 and other small molecular weight GTP binding proteins [112]. TC10 is a member of the Rho family of GTPases, and is also targeted to lipid raft domains as a result of its capacity to undergo posttranslational modification by farnesylation and palmitoylation [113]. Insulin-mediated recruitment of C3G to the lipid raft results in the activation of TC10 via GTP exchange for GDCIP4/2P. TC10 in turn interacts with and recruits Cdc42-interacting protein 4/2 (CIP4/2) to the lipid raft [114]. The signaling events between TC10-CIP4/2 and translocation of GLUT4-containing vesicles have not yet been elucidated. It is tempting to hypothesize that TC10, as a Rho family member, participates in the regulation of the actin cytoskeleton, given that investigators have demonstrated...
a functional requirement for filamentous actin in GLUT4 translocation [115].

Experiments involving transgene expression in adipocytes have provided strong support for the contribution of the CAP/Cbl/TC10 pathway. Deletion of CAP SH3 domains creates a dominant-interfering mutant that blocks binding to Cbl and recruitment of Cbl to lipid rafts, and prevents GLUT4 translocation and stimulation of glucose transport without loss of PI-3 kinase or MAP kinase activation [39]. SoHo deletion mutations in CAP, which prevent binding of the CAP/Cbl complex to flotillin in lipid rafts, also block GLUT4 translocation and glucose transport stimulation [116]. Hyperexpression of C3G potentiates insulin action and shifts the insulin dose–response curve to the left [112]. Overexpression of TC10 (but not other Rho family members) results in miss-targeting to nonlipid raft regions of the plasma membrane, and inhibits recruitment of GLUT4 translocation associated with disruption of cortical actin in adipocytes [113]. Mutant forms of CIP4/2 also inhibit GLUT4 translocation. To illustrate the dependence on caveolin-containing lipid rafts, disruption of rafts by cholesterol-extracting drugs (β-cyclodextrin) or by expression of mutant caveolins blocks insulin stimulation of TC10 and glucose transport [117,118]. These experiments demonstrate that the CAP/Cbl/TC10 pathway is necessary for insulin stimulation of glucose transport. However, it is important to emphasize that these findings are well established in the adipocyte system, whereas, in muscle cells, a definitive role for the CAP/Cbl/TC10 pathway has yet to be firmly established.

In future research, it will be important to elucidate the mechanisms by which activation of Akt/PKB, PKC ζ/λ, and TC10 interact with insulin’s metabolic effector systems and, in particular, with the trafficking and translocation of GLUT4 vesicles (see later). Detailed study of these processes will be necessary to better define defects causing insulin resistance in human disease.

**Muscle contraction**

Acute exercise or muscle contraction results in GLUT4 translocation and stimulates muscle glucose transport activity [119]. This effect occurs without any change in serum insulin concentrations and does not involve activation of PI-3 kinase or Akt/PKB. The mobilization of GLUT4 to the cell surface by acute exercise may involve a different intracellular pool of GLUT4 than that recruited by insulin, and the effects of acute exercise and insulin are partially additive [119]. These observations indicate that signaling systems mediating glucose transport stimulation are different in response to acute exercise versus insulin [120]. This is underscored by the finding that, in muscle from insulin resistant humans and rodents, GLUT4 translocation is impaired in response to insulin but normal in response to acute exercise.

While signal transduction mechanisms are not fully understood, this response to acute exercise appears to be at least partially dependent on increments in intracellular 5′ AMP induced by acute exercise, and subsequent activation of 5′ AMP-activated protein kinase (AMPK) [120–122]. Hypoxia and mitochondrial uncoupling agents also increase 5′ AMP levels and increase cellular glucose transport rates. Investigators in this field have made frequent use of 5-amino-imidazole-4-carboxamide-1-β-D-ribose (AICAR), which acts as a 5′AMP analog by being phosphorylated after cellular uptake and then activating AMPK. AICAR administration mimics acute exercise by stimulating both GLUT4 translocation and glucose transport in a PI-3 kinase independent manner [121,122]. AMPK is an enzyme consisting of one catalytic subunit (α) and two regulatory subunits (β and γ), and exhibits a complex mechanism of regulation that “senses” fuel status in the cell. When ATP levels are low and 5′ AMP is elevated, AMPK activates pathways for ATP regeneration and limits further ATP utilization by modifying activity of multiple metabolic enzymes, including acetyl-CoA carboxylase, hydroxymethyl glutaryl-CoA reductase, creatine kinase, and hormone-sensitive lipase [123]. Subsequent studies using AMPK-inactive mouse models confirmed the importance of this pathway in vivo [124].

**AS160**

Downstream mechanisms leading to translocation of a distinct intracellular GLUT4 pool following muscle contraction have not been fully identified. However, an Akt substrate of 160 kDa (AS160) was identified as part of the signaling network for insulin and Akt leading to GLUT4 translocation [124–126]. Insulin-stimulated phosphorylation of AS160 was demonstrated, and this was sensitive to PI-3 kinase inhibition, placing AS160 in the pathway of insulin-stimulated GLUT4 translocation [126]. In addition, AICAR-stimulated phosphorylation of AS160 was ablated in AMPK-inactive mice [127]. Together, these studies identified AS160 as a point of convergence in insulin-mediated and contraction-mediated GLUT4 translocation. Subsequent investigation revealed that AS160 has a complex pattern of serine/threonine phosphorylation that may define its specificity [128–130].

Subsequently, AS160, designated as gene TBC1D4, was identified as a Rab GTPase activating protein [131] that serves to retain GLUT4 intracellularly, and its activation by either insulin or contraction-mediated signaling releases GLUT4 and allows it to be translocated to the plasma membrane. Rabs are key elements in vesicle trafficking pathways [132–134] that toggle between active and inactive forms through GTP loading and hydrolysis, respectively. The finding that AS160 is a Rab GTPase places it squarely in the machinery of vesicle trafficking. More recently, there is evidence that AS160 plays two roles; one, in which unphosphorylated AS160 restrains GLUT4 translocation, and a second, where phosphorylation of AS160 can facilitate fusion of GLUT4 vesicles with the plasma membrane [131].
Table 12.1 The GLUT gene family

<table>
<thead>
<tr>
<th>Isoform identification*</th>
<th>Class*</th>
<th>HUGO gene nomenclature</th>
<th>Major tissue distribution</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLUT1</td>
<td>I</td>
<td>SLC2A1</td>
<td>Brain microvessels, erythrocytes, placenta, kidney</td>
<td>Ubiquitous, functions as basal transporter, predominates in cultured cell systems</td>
</tr>
<tr>
<td>GLUT2</td>
<td>I</td>
<td>SLC2A2</td>
<td>Liver, kidney, β cell, small intestine</td>
<td>High $K_m$ transporter</td>
</tr>
<tr>
<td>GLUT3</td>
<td>I</td>
<td>SLC2A3</td>
<td>Brain, placenta, fetal muscle</td>
<td>Low $K_m$ transporter, found in tissues metabolically dependent on glucose</td>
</tr>
<tr>
<td>GLUT4</td>
<td>I</td>
<td>SLC2A4</td>
<td>Skeletal muscle, adipocytes, heart</td>
<td>Sequestered intracellularly and translocates to cell surface in response to insulin</td>
</tr>
<tr>
<td>GLUT5</td>
<td>II</td>
<td>SLC2A5</td>
<td>Testes, small intestine</td>
<td>High affinity for fructose</td>
</tr>
<tr>
<td>GLUT6</td>
<td>III</td>
<td>SLC2A6</td>
<td>Spleen, leukocytes, brain</td>
<td>Transports glucose</td>
</tr>
<tr>
<td>GLUT7</td>
<td>II</td>
<td>SLC2A7</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>GLUT8</td>
<td>III</td>
<td>SLC2A8</td>
<td>Testes, blastocyst, brain</td>
<td>Transports glucose, mediates insulin-stimulated glucose transport in blastocyst</td>
</tr>
<tr>
<td>GLUT9</td>
<td>II</td>
<td>SLC2A9</td>
<td>Liver, kidney</td>
<td></td>
</tr>
<tr>
<td>GLUT10</td>
<td>III</td>
<td>SLC2A10</td>
<td>Liver, pancreas</td>
<td></td>
</tr>
<tr>
<td>GLUT11</td>
<td>II</td>
<td>SLC2A11</td>
<td>Heart, muscle</td>
<td>Fructose-inhibitable glucose transport</td>
</tr>
<tr>
<td>GLUT12</td>
<td>III</td>
<td>SLC2A12</td>
<td>Heart, prostate</td>
<td></td>
</tr>
<tr>
<td>GLUT13</td>
<td>III</td>
<td>SLC2A13</td>
<td>Brain</td>
<td>Myoinositol transporter, also referred to as HMIT</td>
</tr>
<tr>
<td>GLUT 14</td>
<td></td>
<td>SLC2A14</td>
<td>Testes</td>
<td>Alternatively spliced long form and short form which is a duplication of GLUT3</td>
</tr>
</tbody>
</table>

*Isoform designation and class conforms to nomenclature proposed by Joost HG et al.: American Journal of Physiology 2002;282:E974–E976. Note: Four untranslated pseudogenes have been reported including a pseudogene previously designated as GLUT6.

The glucose transport effector system

GLUT4 vesicle translocation and trafficking

Glucose transport proteins are the key functional units of the glucose transport effector system. Multiple glucose transporter genes have been identified [66,135] that encode a family of homologous proteins exhibiting different functional properties and marked differences in tissue-specific expression. Members of the GLUT gene family are delineated in Table 12.1. On the basis of hydrophobicity plots, GLUT proteins contain 12 α-helical transmembrane-spanning domains with both the NH$_2$ and COOH terminals extending into the cytoplasm. This model forms six exofacial loops and five endofacial loops. Regarding GLUT1-4 isoforms (class I in Table 12.1), the five endofacial loops include a large hydrophilic central loop, while the first exofacial loop is the largest and contains an asparagine-linked glycosylation site (N57). GLUT4 glycosylation first occurs in the ER, and as GLUT4 travels on to the Golgi apparatus, additional glycan modification can occur in a multistep process, producing multiple versions of GLUT4, distinguishable from each other due to alternative side chain branching or differences in chain length. Alternatively N-glycosylated versions of GLUT4 may explain why GLUT4 differentially migrates during SDS-PAGE separation [136]. A full glycosylation profile of GLUT4 has been shown to be very important for GLUT4 protein stability [136,137] and GLUT4 localization [138]. Thus, GLUT proteins are intimately embedded in membranes, and the most highly conserved regions are the putative membrane-spanning domains that serve a common function, the creation of a pore for facilitative diffusion of monosaccharides.

The major insulin-responsive glucose transporter isoform is GLUT4, which is predominately expressed in insulin target tissues such as skeletal and cardiac muscle and adipose tissue. In the basal state, GLUT4 is predominately localized intracellularly. The major mechanism by which insulin activates glucose transport activity is by recruiting intracellular GLUT4 to the plasma membrane, which results in facilitative diffusion of glucose into the cell where glucose is rapidly phosphorylated by hexokinase and metabolized. Importantly, the glucose transport step is rate-limiting for insulin-stimulated glucose uptake and metabolism in peripheral target tissues under physiologic conditions [135]. Upon dissipation of the insulin signal, deactivation of glucose transport activity is the result of a net reverse translocation of GLUT4 back into the cell interior.

Based on subcellular membrane fractionation experiments, insulin-mediated translocation was initially conceptualized as a two-compartment process with recruitment of GLUT4 from Golgi-enriched microsomal membranes to the plasma membrane. However, these studies consistently indicated that only ~50–60% of microsomal GLUT4 translocated to the plasma membrane fraction, suggesting that there were both insulin responsive and nonresponsive intracellular pools. It is now clear that GLUT4 trafficking involves multiple intracellular membrane compartments [66,105,139,140]. GLUT4 protein is synthesized in the endoplasmic reticulum and travels to the trans-Golgi network (TGN), where GLUT4 is then inserted into
highly specialized vesicles. In basal cells, morphologic studies employing immunohistochemistry and electron microscopy have detected GLUT4 in the perinuclear TGN, in cytosolic tubulovesicular elements, and in subplasma membrane vesicles. In one study, disruption of the TGN using the antibiotic brefeldin A did not impair insulin-stimulated glucose transport activity or surface appearance of GLUT4, ruling out TGN-based GLUT4 as being part of insulin-stimulated glucose uptake [141]. GLUT4 was shown to be largely recruited from small vesicles (GLUT4 storage vesicles, or “GSVs” [142]) located near the endofacial surface of the plasma membrane [141]. Other morphologic data have implicated tubulovesicular structures as the source of the translocating GLUT4 pool [139]. In addition to the GSV pool, GLUT4 also cycles through an early endosomal compartment in an endocytotic/exocytotic pathway (distinct from fluid-phase endocytosis) analogous to that described for other insulin-responsive proteins, such as the mannose-6-phosphate/insulin-like growth factor 2 receptor and the transferrin receptor. GLUT4 in the endosomal recycling system is clearly distinct from the intracellular, insulin-inducible GLUT4 storage compartment [143–147]. For example, ablation of the recycling compartment, accomplished via internalization of a transferrin-horseradish peroxidase conjugate, eliminates vesicles containing transferrin receptors and VAMP3/cellubrevin (endosomal markers) but leaves intact a large fraction of GLUT4-containing vesicles. There are two models for GLUT4 localization post-TGN: (1) the dynamic retention model states GSVs cycle between the specific, insulin-sensitive GSV compartment and the endosomal recycling pool, and (2) the static retention model which proposes GSVs are locked in place until insulin signals the release of the GSV for subsequent membrane translocation. Within both models, GSVs are sequestered intracellularly under basal conditions, and it is the GSV, not the endosomal GLUT4, that represents the major source for insulin-mediated exocytosis and translocation of GLUT4 to the plasma membrane. This unique GSV storage compartment can explain why insulin’s effect to augment cell surface proteins is much greater for GLUT4 (10–40-fold over basal) compared with IGF-1 and transferrin receptors that traffic only in the endosomal recycling pathway.

A model for GLUT4 traffic and translocation is shown in Figure 12.5 based on a consensus of current data. Following early endocytosis, GLUT4 is sorted from other receptor and channel proteins in the recycling pathway and sequestered in a unique storage compartment comprised of GSVs and tubulovesicular elements. Insulin recruits GLUT4 primarily by stimulating exocytotic fusion of GSVs with the plasma membrane resulting in exofacial exposure of functional glucose channels. In basal cells, constitutive cycling of GLUT4 between intracellular and cell surface compartments has been demonstrated which may reflect a proportion of GLUT4 remaining in the endosomal recycling pathway. Insulin increases numbers of plasma membrane GLUT4 primarily by augmenting net exocytosis and also exerts a smaller effect to retard endocytosis. Stimulation of exocytosis proceeds primarily from the GSV pool with a smaller contribution from the constitutive recycling pathway.

**Figure 12.5** Cellular trafficking and itinerary of GLUT4 glucose transporter proteins. The model is based on a consensus of current data, and includes an endosomal recycling pathway, a separate intracellular inducible GLUT4 storage compartment which supplies the majority of GLUT4 recruited by insulin to the cell surface, and an alternative compartment where accumulated GLUT4 does not respond to insulin in insulin-resistant states (see text). (For a color version of this figure, please see color plate section.)
GLUT4 vesicle trafficking is directed by several factors, including targeting motifs intrinsic to GLUT4, posttranslational modifications of GLUT4, and interactions with multiple proteins known to regulate the vectorial flow of cellular vesicles and their cargo proteins. In terms of intrinsic signals, a di-leucine motif at positions 489 and 490 in the cytosolic C-terminal of GLUT4, and an aromatic-based motif (FQQP) in the NH₂ terminus, function as internalization signals for endocytosis [139,148]. While these intrinsic motifs help direct internalization, the signals for intracellular retention in the inducible storage compartment have still not been identified with certainty. Interestingly, upon mutation of the core GLUT4 glycosylation site (within the first exofacial loop at the N terminus), GLUT4 has been shown to accumulate at the cell membrane [138], while at the same time insulin-mediated translocation of intracellular GLUT4 was impaired [137], suggesting GLUT4 glycosylation plays an important role in the trafficking of the GSV. Even so, others have produced contradictory evidence showing that GLUT4 glycosylation does not affect GLUT4 vesicular trafficking [136,137,149]. Some evidence has been obtained suggesting that an acidic cluster of amino acids (TELELYLGP) downstream from the di-leucine motif is important for targeting GLUT4 to the inducible storage compartment [150]. In other experiments, overproduction of the GLUT4 C-terminus resulted in partial GLUT4 translocation and activation of glucose transport, suggesting that this peptide is interacting with an intracellular retention protein [151]. The COOH-terminus di-leucine motif is adjacent to a phosphorylation site (serine at 488) which, when mutated, can impair internalization when expressed in host cells. GLUT4 phosphorylation occurs at several residues within the large cytosolic loop between helices 6 and 7 and also at several sites within the C-terminal cytoplasmic tail. There is a lack of consensus over whether phosphorylation of GLUT4 is increased, decreased, or unaffected upon stimulation with insulin, since evidence has been shown to support all three of these possibilities [152]. Thus, the role of GLUT4 phosphorylation in response to insulin is unclear.

In addition to phosphorylation and N-glycosylation, GLUT4 undergoes ubiquitination (the addition of ubiquitin, a 76-amino acid protein) at residues localized throughout the intracellular loops of the protein, and this modification is important for GLUT4 to properly localize to the GSV [153]. GLUT4 also undergoes SUMOylation, which is the addition of the 101-amino acid protein SUMO-1, although the exact location of SUMOylation on GLUT4 is unknown. SUMOylation of GLUT4 is mediated by ubiquitin conjugating enzyme 9 (Ubc9), and was shown to considerably extend the half-life of GLUT4 protein expression, suggesting that this modification is crucial for GLUT4 protein stability [154,155]. Palmitoylation, the covalent attachment of fatty acids, of GLUT4 has also been observed [156], although the role of this posttranslational modification is unclear.

GSV traffic is controlled by multiple protein subclasses, also known to be involved in the regulated endocytosis/exocytosis of synaptic and neurosecretory vesicles [135,139,140,143–148]. Multiple trafficking proteins are actual constituents of the GLUT4 vesicle; GLUT4 vesicle cargo proteins include vesicle-associated membrane protein-2 (VAMP-2) or synaptobrevin, VAMP-3/cellubrevin, syntaxin 4, 35-39 kDa secretory carrier-associated membrane proteins (SCAMPs), cellugyrin, acyl-CoA synthetase-1, and small molecular weight GTP binding proteins in the Rab and Arf subclasses. Of these proteins, VAMP-2, VAMP-3/cellubrevin, SCAMPs, and cellugyrin are known to be constituents of GLUT4-containing vesicles in the endosomal recycling pathway where they co-localize with IGF-1 and transferrin receptors. Of these, only VAMP-2 is present together with GLUT4 in the intracellular storage compartment. Several other constituents of the GSV are an aminopeptidase, referred to as the insulin-regulated aminopeptidase (IRAP) [157], low-density lipoprotein receptor-related protein 1 (LRP1), and sortilin, all of which contribute to promoting GLUT4 localization to the GSV [158–160]. These proteins translocate with GLUT4, while IRAP in particular has commonly been used to track the cellular itinerary of GSV's which is translocated along with GLUT4 to the plasma membrane.

The v-SNARE/t-SNARE model developed by Rothman and others explains how vesicles traffic to specific membrane compartments [161]. In this model, shown in Figure 12.6, a docking and fusion reaction is initiated via the high affinity interaction between a ligand on the transport vesicle (v-SNARE) and a receptor in the target membrane (t-SNARE). In the case of GLUT4 vesicle translocation from the inducible storage compartment to the cell surface, evidence suggests that VAMP-2 functions as the v-SNARE and syntaxin 4 as the plasma membrane t-SNARE [135,139,140,143–148]. Other proteins facilitate and regulate this interaction. The N-ethylmaleimide sensitive factor (NSF) activates the v-SNARE for docking competency, and three proteins, the soluble NSF attachment 23kDa protein (SNAP23), the mouse Unc homologue 18c (Munc-18c), and the syntaxin4-interacting protein (Synip), regulate docking efficiency at the level of the t-SNARE by facilitating or blocking access to syntaxin4 [162]. While SNAP23, Munc-18c, and Synip form a complex with syntaxin4 at the endofacial plasma membrane, their precise roles in modulating interaction between VAMP-2 and syntaxin4 are not fully understood. Synip appears to block access of VAMP-2 to syntaxin4 under basal conditions, and is released from the complex following insulin stimulation allowing docking of v-SNARE (VAMP-2) to the t-SNARE (syntaxin4). The function of Munc-18c appears to be more complex with evidence that it can both inhibit (basal conditions) and facilitate (insulin-stimulated condition) VAMP-2-syntaxin4 interaction. The Rab family of monomeric GTPases also play an important role in vesicle traffic; different isoforms are targeted to specific organelle membranes and direct the vectorial flow of vesicle proteins from one compartment to another [163,164]. Rab GTPases also appear to catalyze the union of v- and t-SNAREs in the docking process. Rab4 has
been demonstrated to be associated with GLUT4 vesicles in basal cells, but is redistributed to cytosol in response to insulin where it associates with its GDP-association inhibitor [165]. Furthermore, Rab4 has been shown to interact with syntaxin4 in vitro, and overexpression of Rab4 or a COOH-terminal domain peptide, prevented insulin-stimulated GLUT4 translocation in adipocytes [166]. Undoubtedly, there are additional factors participating in GLUT4 vesicle targeting, docking, and fusion in insulin target tissues that remain undiscovered.

The cellular cytoskeleton

There is increasing recognition that GLUT4 translocation involves the actin cytoskeleton [66,167–169]. One line of evidence supporting this hypothesis is that disruption of the actin cytoskeleton by cytochalasin D or latrunculin also inhibits insulin-mediated GLUT4 translocation [167,168]. In addition, insulin stimulates cytoskeletal rearrangement, in particular the cortical actin fibers subtending the plasma membrane, and regulates molecules that control actin polymerization. Insulin-stimulated actin reorganization is dependent upon members of the small G-proteins of the Rho family [66,115], with the protein Rac1 active in muscle and both TC10 and Cdc42 in adipose tissue. Insulin stimulates GTP loading of Rac1, TC10, and Cdc42 followed by activation of the nucleation-promoting factor (NPF) of the Wiskott–Aldrich syndrome protein (WASP) family member, N-WASP [169,170]. The WASP proteins are responsible for the binding and activation of the Arp2/3 complex. Arp2/3 then acts as the true actin nucleator by binding to an existing actin filament to create new actin branched networks. It is attractive to hypothesize that activation of TC10 via the CAP/Cbl/TC10 pathway could mediate regulation of the actin cytoskeleton in response to insulin, since TC10 and other Rho GTPases are generally known to affect biochemical modulators of cytoskeletal function [171]. However, PI-3 kinase can also affect cortical actin polymerization and events such as membrane ruffling and filopodia formation; therefore, other insulin signaling pathways, like Rac1, could also contribute to this biologic effect. The exact role of the actin cytoskeleton has not yet been delineated; however, one proposal is that actin filaments serve as a scaffold to direct GLUT4 vesicle trafficking using actin-based motility as a motor for vesicle movement.

The microtubule cytoskeleton is also involved in GLUT4 vesicle trafficking [140,172,173]. The GSV translocate along microtubule tracks to arrive at the inducible storage compartment proximal to the plasma membrane [174,175], and it is not surprising that microtubule proteins such as dynein and kinesin have been co-purified with GLUT4. Disruption of microtubules using depolymerizing agents inhibits GLUT4
translocation and glucose transport stimulation, and expression of a dominant-interfering dynein mutant was shown to inhibit GLUT4 endocytosis in a process that may require Rab5 [176]. This transport has been shown to be mediated by the cargo transport motor proteins of the kinesin family [177] including KIF5B [172] and KIF3 [178]. Upon receiving the insulin signal, tether containing UBX domain for GLUT4 (TUG) was found to disassociate from GLUT4, which has been postulated to act as the release of a “handbrake mechanism,” allowing for GLUT4 to travel to the plasma membrane [179]. Upon arrival at the interior surface of the cell, the GSV undergoes the “docking” phase of translocation, through association with the exocyst complex, which has been proposed to function in the targeting of the GSV to areas on the plasma membrane rich in the proteins involved in fusion of the GSV with the membrane [180,181]. The GSV then undergoes the transition to the docking/fusion phase, which is mediated by the SNARE complex.

With the advent of real-time total internal reflection fluorescence microscopy (TIRFM), which visualizes just the first 200 nm of the intracellular surface of the cell membrane, it was confirmed that microtubules form a network directly underneath the entire plasma membrane, which was proposed to allow GSVs to scan the cytoplasmic surface of the cell membrane during the highly mobile basal state [182]. TIRFM has also shown that defective actin remodeling disrupted the exocytic fusion step of the GSV, whereas the insulin-stimulated increase of GLUT4 vesicles proximal to the plasma membrane was not affected [183]. TIRFM performed in living 3T3-L1 adipocytes proved that insulin-stimulated fusion of GSVs occurs proximal to microtubules at the plasma membrane. This study also captured, for the first time, video evidence in living adipocytes of the substantial increase of microtubule density and curvature that occurs at the plasma membrane during insulin stimulation. The authors did find long-range movements of GSVs along microtubules, although the GSVs rarely moved long distances before fusion, suggesting that long-range GSV trafficking is not involved in insulin-stimulated GLUT4 insertion into the membrane. Nocodazole treatment, for example, did not significantly decrease the number of GSV fusion events stimulated by insulin, leading the authors to conclude that microtubules are not vital for the insertion of GSVs into the plasma membrane, but rather play a more important role in site selection for delivery of GLUT4 prior to fusion [184]. In line with this hypothesis, CLASP2, a microtubule associated protein that targets microtubules to the interior surface of the cell cortex, was found to undergo insulin stimulated co-localization with GLUT4 at the plasma membrane. This led the authors to propose that the distal end of microtubules undergo CLASP2 targeting to landing zones on the cell cortex, thereby situating microtubule-bound GSVs proximal to the plasma membrane [185]. An attractive hypothesis is that GLUT4 begins exocytosis in a microtubular compartment, and then transfers to actin scaffolds that connect the microtubule cytoskeleton with the plasma membrane, possibly through the protein actinin-4 [186], which then positions GLUT4 vesicles for docking and membrane fusion. Upon endocytosis, GLUT4 traffics in an endosomal compartment regulated by actin filaments and eventually is sorted back to the microtubular inducible storage compartment. While this model is compelling, additional research is necessary to define the role of cytoskeletal elements in GLUT4 translocation. In any case, these considerations emphasize the complex nature of the glucose transport effector system, which functionally incorporates both vesicle traffic and cytoskeletal systems.

**Insulin resistance in humans: abnormalities in insulin signaling and in the glucose transport effector system**

**Abnormalities in insulin signaling**

Although the key elements of the insulin signaling network have been defined by studies employing molecular and cell biology techniques, a number of studies have focused on translating this knowledge to human studies in a clinical research setting. Insulin resistance in peripheral tissues characterizes obesity and T2DM, and is involved in the pathogenesis of diabetes [187]. Studies using hyperinsulinemic, euglycemic clamps have demonstrated that systemic insulin resistance is associated with reduced stimulation of glucose transport in muscle and adipose tissues biopsied from these same subjects [188]. Further studies in these insulin target tissues have elucidated defects in insulin signaling as well as defects intrinsic to the glucose transport system in insulin-resistant humans.

Regarding insulin signal transduction, investigators have demonstrated reductions in insulin-stimulated tyrosine phosphorylation of both insulin receptor and IRS-1, decreased association of PI-3 kinase with IRS-1, and more modest decrements in insulin-stimulated Akt phosphorylation [189,190]. Despite these defects, insulin stimulation of ERK2 phosphorylation was normal or even elevated in muscle from patients with T2DM [189,190]. This dichotomy between insulin-stimulated PI-3 kinase and MAP kinase signaling later was extended to insulin signaling in vascular cells under conditions that produce insulin resistance [191]. The implications of this dichotomy remain unclear, but it is conceivable that maintenance of this imbalance could lead to worsening of insulin resistance through serine phosphorylation of IRS-1 by ERK [192]. To probe causal relationships, other studies employed perturbations to either enhance or diminish in vivo insulin sensitivity to determine if there were commensurate changes in phosphorylation/dephosphorylation of the insulin receptor, IRS-1, and PI-3 kinase. Using muscle biopsies and euglycemic clamps, investigators have shown beneficial effects of treatment with thiazolidinediones [193], exercise [194], caloric restriction [195], and tight glucose control [196]. In contrast, insulin resistance provoking treatments like infusion of a triglyceride emulsion worsened insulin signaling defects [197].
One of the methodological breakthroughs that enabled investigators to define mechanisms of phosphorylation-induced activation or inactivation of insulin signaling proteins was the use of motif-specific antibodies directed against phosphopeptide motifs present in these proteins. Although this provided a powerful and facile method approach in \textit{in vivo} studies, there always remain questions of specificity and quantification. Furthermore, with these techniques, only changes in known phosphorylation events can be assessed. A relatively new addition to the tool kit used to analyze insulin signaling is mass spectrometry based proteomics techniques, especially applied to posttranslational modifications, and in particular, phosphorylation, due to its importance in regulating activity of signaling proteins. Mass spectrometry allows for identification of novel phosphorylation sites in known insulin signaling proteins and the discovery of new insulin signaling proteins that exhibit insulin-induced changes in phosphorylation. These techniques have proven to be invaluable for deciphering the complexity of serine and threonine phosphorylation sites in IRS proteins [198,199] and AS160 [130]. These methods allow for simultaneous quantification of many phosphorylation sites at the same time, unlike immunoblot analysis. For example, Humphries et al. were able to identify 37,248 phosphorylation sites on 5705 proteins in 3T3-L1 adipocytes, with approximately 15% responding to insulin. This led to the discovery that SIN1, a core component of the mTORC2 complex, is an Akt substrate, and the phosphorylation of SIN1 by Akt regulates mTORC2 activity in response to growth factors [200].

The glucose transport system

Given that plasma membrane glucose transport is rate-limiting for insulin-stimulated glucose metabolism in insulin target cells, defects in GLUT4 expression, diminished functional activity, or impaired translocation could readily explain insulin resistance [201]. The multicompartmental nature of GLUT4 cellular trafficking, requiring a multiplicity of regulatory proteins, indicates that there are many potential sites for defects that could impair GLUT4 translocation. Indeed, defects intrinsic to the glucose transport system have been shown to be an important cause of insulin resistance in muscle and fat tissues from insulin-resistant and diabetic patients. In adipocytes, cellular depletion of GLUT4 transporters is a major mechanism of insulin resistance in obesity and T2DM [202], while GLUT4 expression is relatively normal in skeletal muscle [203]. However, in both muscle and fat, GLUT4 accumulates or is redistributed to a dense membrane compartment under basal conditions, and this abnormality is linked to impaired GLUT4 translocation to the plasma membrane in response to insulin [204–206] (see Figure 12.6). These data are indicative of a trafficking or targeting abnormality impairing GLUT4 translocation in tissues derived from insulin-resistant subjects. Additional study of pathways and factors involved in abnormal GLUT4 traffic may elucidate mechanisms causing human insulin resistance and establish targets for drug development.

**Modulation of insulin action**

Important recent advances in the study of insulin signaling include the elucidation of multiple mechanisms for dampening or inhibiting insulin signal transduction. The studies have provided insight into the extensive regulation of insulin action, and have also identified potential therapeutic targets, since blocking these inhibitory mechanisms could enhance insulin sensitivity. These mechanisms are summarized in Figure 12.7.

**Serine-threonine phosphorylation of the insulin receptor and IRS proteins**

Serine-threonine phosphorylation of insulin receptors and insulin substrate docking proteins has emerged as a major mechanism for modulation of insulin signal transduction. Serine phosphorylation of the insulin receptor diminishes its tyrosine kinase activity [10]. Serine phosphorylation of IRS-1 (and other IRSs) decreases receptor-IRS coupling by inhibiting insulin-mediated receptor phosphorylation, tyrosine phosphorylation of IRS-1, binding and activation of PI-3 kinase, and stimulation of glucose transport [207–209]. There are consensus sequences in IRS-1 that make it susceptible to a wide variety of serine-threonine kinases including PKC, PKA, Akt/PKB, MAP kinase, GSK3, casein kinase II, Cdc2 kinase, and JNK. Several of these kinases have been shown to function as physiologic modulators causing desensitization of insulin signaling pathways [210].

**Protein kinase C**

Protein kinase C isoforms are categorized as classic (α,β,γ), novel (δ,θ,ε,η,μ), and atypical (ζ,λ,ι) depending on their ability to be activated by calcium and diacylglycerol (DAG). These serine-threonine kinases can act on multiple substrates, including IRS docking proteins and the insulin receptor [10,207,208]. Serine-threonine phosphorylation of IRS impairs its ability to associate with the insulin receptor and with PI-3 kinase, resulting in desensitization of the PI-3 kinase pathway. Hyperinsulinemia, hyperglycemia, and elevated circulating free fatty acids increase intracellular DAG, which in turn activates conventional and novel PKC isoforms principally via recruitment to the plasma membrane. These conditions are associated with increased serine-threonine phosphorylation and diminished function of insulin receptors and IRS proteins. In addition, insulin activates atypical PKCs, such as PKCζ, via the PI-3 kinase pathway, and atypical PKCs are then also capable of phosphorylating and desensitizing IRS [211,212]. Thus, PKCζ could be involved both in direct promulgation (Figure 12.2) and in feedback inhibition (Figure 12.7) of insulin signal transduction.

**Tumor necrosis-α (TNFα)**

TNFα is a cytokine produced by immune cells and also by adipocytes and muscle tissue. While having little impact on systemic circulating concentrations, TNFα expression is increased...
Mechanisms of insulin signal transduction

Figure 12.7  Mechanisms for modulation of insulin signal transduction. The figure illustrates pathways resulting in the desensitization of insulin signal transduction, including serine/threonine phosphorylation of the insulin receptor or IRS by PKC, IKKβ, or JNK; inhibition of the insulin receptor tyrosine kinase by PC-1; tyrosine phosphatase action of PTPase-1B and LAR; catabolism of 3′-phosphoinositide moieties by SHIP-2 or PTEN; inhibition of Akt/PKB activation by ceramide; and interference with substrate binding (STAT5B) to the insulin receptor β subunit by SOCS-3. Abbreviations: PC-1, plasma cell membrane glycoprotein PC-1; SOCS-3, suppressor of cytokine signaling 3; STAT-5B, signal transducer and activator of transcription 5B; PTPase-1B, protein tyrosine phosphatase 1B; LAR, leukocyte antigen-related tyrosine phosphatase; cPKC, classic protein kinase C isoforms; nPKC, novel protein kinase C isoforms; PKC-ζ, protein kinase Cζ; JNK, c-Jun N-terminal kinase; IKK-β, kinase of inhibitor of kappalight chaingene enhancer in B cells β; TNF-α, tumor necrosis factor α; IRS, insulin receptor substrate protein; ser/thr, a serine or threonine residue; Y, tyrosine residue; PI-3 kinase, phosphatidylinositide-3 kinase;SHIP-2, SH2-containing inositolphosphatase2; PTEN, phosphataseandtensinhomologue; Akt/PKB, proteinkinaseB. (For a color version of this figure, please see color plate section.)

in adipose and muscle tissues as a function of insulin resistance [213,214], raising the possibility that the cytokine could be inducing cellular insulin resistance via autocrine/paracrine effects. Consistent with this idea, TNFα has been shown to cause insulin resistance in cultured cell models. While the mechanisms underlying desensitization have not been fully elucidated, TNFα leads to an increase in serine phosphorylation of IRS-1, thereby decreasing both insulin receptor tyrosine kinase activity and downstream insulin signal transduction [215–217]. One possible branch of the TNFα signaling pathway that could mediate IRS phosphorylation results is the activation of the stress-induced kinase, c-JUN NH2-terminal kinase (JNK). JNK is a serine kinase with many intracellular substrates including insulin receptor substrate docking proteins, IRS-1, IRS-2, Shc, and Gab-1. These docking molecules contain JNK binding motifs that facilitate JNK-mediated phosphorylation of serine residues. In the case of IRS-1, JNK action results in the phosphorylation of a serine residue (ser307 in mice or ser312 in humans) that lies within the PTB domain; phosphorylation at this site interferes with PTB function reducing interaction between insulin receptor and IRS-1 [218]. Another potential mechanism involves insulin stimulation of sphingomyelinase activity causing release of ceramide from sphingomyelin in lipid raft domains [81,219]. Exogenous ceramide induces insulin resistance in cultured adipocytes associated with decreased Akt/PKB phosphorylation and GLUT4 translocation. Regardless of underlying mechanisms, abrogation of TNFα action by targeted disruption of TNFα receptor isoforms (p55 and p75) results in enhanced insulin sensitivity in mice [219].

IkB kinase-β (IKKβ)

Treatment with high doses of salicylates improves glucose tolerance and enhances insulin sensitivity in humans and rodents [220]. An important biologic effect of salicylates is the inhibition of the IKKβ. Because IKKβ is a serine kinase, these observations implicate this serine kinase cascade in desensitization of the insulin signal, perhaps through serine phosphorylation of IRS or the insulin receptor. Predictably, salicylates would then enhance insulin sensitivity by causing a subsequent decrease in IKKβ-mediated serine phosphorylation
of IRS. Consistent with this hypothesis, heterozygous ablation of the IKKβ gene in mice prevents insulin resistance induced by high-fat feeding or lipid-infusion, and in obese leptin-deficient ob/ob mice [221,222]. IKKβ can also activate nuclear transcription factor kappa-B (NF-κB) via phosphorylation of 1κB, a physically bound inhibitor of NF-κB. Upon phosphorylation, 1κB dissociates from NF-κB allowing NF-κB to enter the nucleus where it can activate multiple genes involved in inflammation, immunity, and the stress response, including TNFα. Thus, inhibition of IKKβ could also enhance insulin signaling by blocking activation of NF-κB.

**Protein phosphotyrosine phosphatases**

Endogenous phosphotyrosine phosphatases (PTPases) are able to dephosphorylate tyrosine residues on the insulin receptor β subunit and insulin receptor substrate docking molecules, resulting in a dampening of insulin signal transduction [223]. Membrane-associated PTPase activity is increased in skeletal muscle from patients with T2DM [224], principally due to increments in cytosolic PTPase-1B and membrane-associated leukocyte antigen-related phosphatase (LAR) [225,226]. The ability of PTPases to modulate insulin signaling has been demonstrated in mice with genetic ablation of PTPase-1B, which exhibit enhanced insulin sensitivity, increased insulin-mediated tyrosine phosphorylation of the receptor and IRS, and a failure to develop insulin resistance when fed a high-fat diet [227].

The data demonstrate that PTPase-1B is able to negatively modulate insulin signaling through dephosphorylation of tyrosine residues in the insulin receptor and IRSs. Inhibition of PTPase action represents a clinically attractive therapeutic target for increasing insulin sensitivity.

**Other mechanisms**

**Suppressor of cytokine signaling-3**

Observations pertaining to suppressor of cytokine signaling-3 (SOCS-3) point to a new mechanism for insulin signal modulation [41]. The SOCS family of proteins (CIS and SOCS 1–7) were originally described as a negative feedback loop for cytokine receptor signaling through Janus kinase (JαK). JαK phosphorylates and activates signal transducer and activator of transcription (STAT), which is a transcription factor that augments SOCS gene expression. SOCS proteins then feedback to inhibit tyrosine-phosphorylated cytokine receptors, via either competitive binding through their SH2 domain preventing phosphorylation of cytokine receptor substrates or by binding and inhibiting Jak tyrosine kinases. Evidence suggests that a similar mechanism may be operative for the insulin receptor [228]. Insulin stimulates STAT5B activation by direct tyrosine phosphorylation after STAT5B binds to insulin receptor phosphotyrosine 972 via its SH2 domain (i.e., this interaction does not involve a PTB as is the case with IRS). STAT5B induces expression of SOCS-3, which then competes with STAT5B binding to the insulin receptor inhibiting further STAT5B activation. Thus, SOCS-3 could function as a negative regulator for insulin signaling, not only with respect to STAT5B activation, but also for other signaling pathways to the extent that SOCS-3 could prevent the interaction between the PTB domain of IRS and the insulin receptor. It is intriguing that several factors that induce cellular insulin resistance also induce SOCS-3 expression, including TNFα [229], growth hormone [230], and leptin [231].

**Plasma differentiation factor-1 (PC-1)**

PC-1 is a membrane glycoprotein with pyrophosphatase activity that appears to act as an intrinsic inhibitor of the insulin receptor tyrosine kinase through an undefined mechanism [232–234]. Muscle expression of PC-1 is elevated in diabetes and obesity, and correlates with diminished insulin receptor tyrosine phosphorylation and muscle glucose uptake [235].

**Regulation of gene transcription**

**Insulin response elements**

Insulin regulates the expression of many individual proteins through effects on translation, mRNA stability, and gene transcription. The stimulatory effect on translation generally results from increases in the initial phase of translation of mRNA by ribosomes, and is discussed in Chapter 16: Insulin regulation of protein metabolism. Regulation of gene transcription can be due to effects on transcription factors that are directly modified via insulin signaling mechanisms, or that are indirect as the result of stimulation of glucose metabolism. For example, genes encoding pyruvate kinase, fatty acid synthase, acetyl-CoA carboxylase, and stearoyl-CoA desaturase require increases in both insulin and glucose for induction, and PEPCK expression can be suppressed independently by insulin or glucose. The promoter elements (cis acting sequences) that mediate changes in gene transcription are referred to as insulin response sequences (IRS) or insulin response elements (IRE). There clearly exist multiple distinct classes of IREs that participate in the regulation of different gene promoters, rather than a single consensus nucleotide sequence [236]. These IREs can mediate either positive effects (GAPDH, pyruvate kinase, fatty acid synthase, somatostatin, c-fos, glucokinase, apoA1) or negative effects (e.g. PEPCK, glucose-6 phosphatase, tyrosine aminotransferase, apoCIII, glucagon) on gene transcription. It is also important to consider the identity of the transcription factors (trans acting factors) activated by insulin that bind to the IREs and alter gene transcription rates. While the identification of transcription factors has progressed more slowly than with IREs, there has been progress in identifying transcription factors and signaling mechanisms involved in insulin regulation of specific genes.

Insulin regulation of gene expression has been most readily attributed to the mitogenic MAP kinase pathway. However, insulin inhibition of PEPCK, IGFBP-1, and glucose-6-phosphatase genes employs the PEPCK-like IRE with consensus
sequence T(G/A)TT(T/G)(G/T) in the core motif, and this occurs through a PI-3 kinase-dependent pathway [236]. The downstream targets of PI-3 kinase are somewhat controversial with studies arguing for and against the involvement of Akt/PKB and a recent study supporting a role for GSK-3 in the inhibition of PEPCK and glucose-6-phosphatase genes [237]. Nevertheless, PI-3 kinase pathways can also regulate gene transcription, and this calls into question notions of distinct and separate metabolic and mitogenic insulin action pathways in the web of insulin signal transduction.

**Tissue and serum response elements**

The ability of insulin to alter gene expression via the mitogenic MAP kinase pathway has long been recognized [238]. This classic mechanism involves phosphorylation of transcription factors in the activating protein-1 (AP-1) family such as c-Jun and c-fos which interact with the TPA response element (TRE), and ternary complex factors such as Elk-1, SAP-1, and SAP-2 which interact with the serum response element (SRE). Both elements can be found in single gene promoters, for example, in the c-fos gene which is induced by insulin through activation of both c-Jun and Elk-1.

**Forkhead transcription factors**

Genetic studies in the nematode, Caenorhabditis elegans, have led to the discovery of an additional pathway for insulin-regulated gene expression [239]. One of the developmental stages in C. elegans is the dauer stage characterized by increased longevity, reduced metabolic activity, and increased body fat. A series of specific gene mutations, termed Daf alleles, has defined a signal transduction pathway leading to a constitutive dauer phenotype. Daf genes include the homologues of the insulin/IGF-1 receptor, PI-3 kinase, PDK-1, and Akt/PKB isoform genes in mammals, as well as Daf-16 which encodes a transcription factor with homology to the mammalian forkhead family of transcription factors. Forkhead transcription factors were first identified as being involved in the forkhead mutation in Drosophila, and the subclass of mammalian forkhead proteins with greatest homology to Daf-16 was first identified in alveolar rhabdomyosarcoma; this latter subclass is referred to as forkhead in human rhabdomyosarcoma (FKHR). The FKHR family is comprised of three expressed genes, FKHR or FOXO1, FKHR L1 or FOXO3, and AFX or FOXO4 [240]. FOXO1 is the most highly expressed FKHR isoform in insulin-responsive tissues such as liver, adipose tissue, and pancreatic β cells.

In nematodes and in mammals, the FKHR proteins represent an important signal transduction pathway by which insulin receptors modulate gene expression. Insulin stimulation results in PI-3 kinase-dependent phosphorylation of FKHR transcription factors by Akt/PKB at three different phosphorylation sites (Thr-24, Ser-256, and Ser-319 in FOXO1) [240]. Under basal conditions, FKHR proteins reside within the nucleus, but, upon phosphorylation, dissociate from their DNA binding site, and are then excluded from the nucleus and retained in the cytosol. Replacement of Thr-24 or Ser-256 by alanine results in loss of phosphorylation by Akt/PKB, lack of FOXO1 nuclear export, and failure of insulin-mediated promoter suppression [241]. The relocation of FKHR out of the nucleus represents an effective mechanism for insulin in the suppression of transcriptional activity. For example, FKHR proteins bind to cis elements in the promoters of the PEPCK, glucose-6-phosphatase, and IGFBP-1 genes and augment transcriptional activity. Transactivation by FKHR is prevented when cells are stimulated by insulin [241–243]. The effects on PEPCK and glucose-6-phosphatase genes have led some investigators to the conclusion that FKHR could help coordinate the overall suppression of gluconeogenesis and hepatic glucose production by insulin. However, FKHR proteins are not involved in the regulation of all genes that are suppressed by insulin. Furthermore, even in the case of PEPCK and glucose-6-phosphatase, FKHR cannot by itself account for all of transcriptional inhibition under physiologic conditions. FKHR can clearly mediate transactivation and insulin suppression of glucose-6-phosphatase and PEPCK when FKHR is overexpressed in cells; however, at physiologic levels, FKHR may not be essential for gene regulation, suggesting that other transcription factors predominate in the regulation of these two genes [243,244].

Another transcription factor, peroxisome proliferative activated receptor-γ co-activator 1 (PGC-1), was first identified as a factor involved in brown fat adipogenesis [245]. PGC-1 also plays an important role in the regulated expression of gluconeogenic enzymes, even at physiologic concentrations within hepatocytes [246,247]. In addition, PGC-1 is induced in the liver by glucagon and glucocorticoids in the context of fasting, insulin deficiency, and diabetes, and then participates in the induction of the gluconeogenic program. However, in regulating hepatic gluconeogenesis, PGC-1 may physically interact with FOXO1 and function as its co-activator [248]. In this role, FOXO1 function appears to be necessary for full induction of gluconeogenic genes by PGC-1 in hepatic cells and mouse liver. Several observations indicate that insulin may suppress gluconeogenic genes by interfering with the interaction between PGC-1 and FOXO1. For example, Akt/PKB is able to phosphorylate FOXO1, and this prevents the binding and co-activation by PGC-1. Also, insulin is able to suppress the induction of gluconeogenic genes by PGC-1 but not in the presence of a constitutively active FOXO1 mutant that is insensitive to insulin [248]. These studies indicate that a physiologic role for FOXO1 in regulating gluconeogenic gene expression may only become evident when this is examined in the context of co-activation by PGC-1.

FKHR may also play other import roles in the biology of insulin-responsive cells. FOXO1 is involved in the stimulation of pancreatic β-cell proliferation and regulation of Pdx1 expression by insulin [249]. FOXO1 also appears to contribute to the complex coordination of transcriptional events involved in adipocyte differentiation. Expression of a constitutively-active FOXO1 mutant prevents differentiation of pre-adipocytes, and
FOXO1 haploinsufficiency in mice protects from diet-induced glucose intolerance and hyperinsulinemia [250].

**Sterol response element binding protein-1c (SREBP-1c)**

Another exciting observation is the role of SREBP-1c in insulin-regulated gene expression. The sterol response element binding protein family of transcription factors is classically viewed as being involved in the regulation of genes in response to the cellular availability of cholesterol [251]. This family includes SREBP-1a and SREBP-1c, which are encoded by a single gene and differ only in their first exon through use of alternative transcription start sites, and the homologous SREBP-2 encoded by a second gene [252]. Evidence has accumulated that SREBP-1c is regulated primarily by insulin rather than cholesterol availability, and is also involved in adipocyte differentiation which explains the factor’s alternative designation as adipocyte differentiation and differentiation factor-1 (ADD1) [253]. SREBP-1c/ADD1 is most highly expressed in liver, white adipose tissue, muscle, adrenal gland, and brain. SREBPs contain an NH2-terminal domain consisting of a basic helix-loop-helix leucine zipper transcription factor, a central domain comprised of two transmembrane-spanning regions, and a COOH-terminal regulatory domain. In this form, SREBPs are associated with membranes in the endoplasmic reticulum. When membrane cholesterol concentrations are decreased, SREBP-1a and SREBP-2 undergo two sequential proteolytic cleavages, releasing the transcriptional domain, and allowing for its translocation into the nucleus [254]. These transcription factors then bind sterol responsive elements (SREs) in specific gene promoters and induce genes involved in cholesterol biosynthesis (i.e., HMG-CoA reductase, HMG-CoA synthase) and the LDL receptor, resulting in the cellular replenishment of cholesterol. Primarily regulation of SREBP-1c by insulin was first manifested in livers of fasted rodents re-fed with a high carbohydrate diet [255–257]. Subsequently, insulin was demonstrated to stimulate the expression and transcription of SREBP-1c in cell lines and in adipose, liver, and muscle tissues. This insulin effect is mediated through IRS-1 and PI-3 kinase, and possibly Akt/PKB [258], although downstream targets and transcription factors mediating induction of SREBP-1c gene transcription have not been elucidated. Regardless, SREBP-1c appears to play an important role in the regulation of specific genes in response to insulin. In experiments involving overexpression of wild-type, constitutively active, and dominant interfering mutants, SREBP-1c has been shown to be involved in the induction of fatty acid synthase and leptin in adipose cells, and glucokinase, pyruvate kinase, fatty acid synthase, and acetyl-CoA carboxylase in the liver [259]. Thus, SREBP-1c mediates a positive transcriptional effect on genes promoting glycolysis and lipogenesis, and seems to prepare the liver for carbohydrate availability following a meal. While containing SREs in their promoter regions, the above-mentioned genes are regulated by multiple hormones as well as metabolic substrates, and the relative contribution of SREBP-1c in overall transactivation by insulin has not been defined.

**Signal transducer and activator of transcription 5B (STAT5B)**

As discussed earlier in relationship to SOCS-3, the insulin receptor directly activates the STAT5B transcription factor via tyrosine phosphorylation [41]. STAT5B binds phosphorytose 972 of the insulin receptor, which is the same juxtamembrane phosphorytose involved in IRS binding via their PTB domains. Binding of STAT5B is the result of a SH2 domain interaction as opposed to the NPXpY-consensus sequence characterizing PTB domain interactions. Direct tyrosine phosphorylation of STAT5B by the insulin receptor kinase allows for dimerization through SH2 domain interactions and translocation into the nucleus [260,261]. This is in contrast to the general mechanism of STAT factor activation by cytokine receptors, which employs Jak kinase as the activating kinase. The complement of genes regulated through insulin activation of STAT5B has not been extensively studied. However, both glucokinase and SOCS-3 genes have STAT5 binding sites in their promoter regions, and transcriptional activity is stimulated by insulin via STAT5B [41,261].

**Acknowledgments**

The work was supported by research grants from the National Institutes of Health [DK-38765 (WTG), DK-083562 (WTG), DK047936 (LJM), DK066483 (LJM)], the UAB Diabetes Research Center (P60-DK079626), the American Diabetes Association, and the Merit Review program of the Department of Veterans Affairs.

**References**

7. Frattali AL, Treadway JL, Pessin JE: Transmembrane signaling by the human insulin receptor kinase. Relationship between intramolecular beta subunit trans- and cis-autophosphorylation...


25 O’Rahilly S, Moller D: Mutant insulin receptors in syndromes of insulin resistance. *Clinical Endocrinology (Oxf)* **1992;36:**121 – 132.


146 Malide D, Dwyer NK, Blanchette-Mackie EJ, Cushman SW: Immunocytochemical evidence that Glut4 resides in a specialized


CHAPTER 13

Regulation of glucose metabolism in liver

Shawn C. Burgess
Division of Metabolic Mechanisms of Disease and Department of Pharmacology, University of Texas Southwestern Medical Center, Dallas, TX, USA

Key points
- Liver produces and consumes glucose.
- Hepatic glucose production is comprised of glycogenolysis and gluconeogenesis.
- Hepatic glucose utilization replenishes glycogen and provides substrate for lipogenesis.
- Hepatic glucose metabolism is regulated by both transcriptional and metabolic mechanisms.

Introduction
The liver regulates nutrient supply to the entire body. Antoine Lavoisier first scientifically disputed the concept of vitalism by proving that the combustion of material in the blood formed the basis of life and mechanical work by muscle. He suggested that these nutrients were not only absorbed by digestion but that liver played an important anabolic role in maintaining life [2]. Less than a century later, Claude Bernard observed that blood glucose persists in the absence of food, and that it is produced by liver from a substance that he called glycogen. Others pointed out that glucose is continuously supplied to the vessels of even the fasted (glycogen depleted) liver, which may have been the earliest detection of gluconeogenesis. Based on lower glucose in the venous blood of muscle, Bernard proposed that glucose was one of the combustible nutrients suggested by Lavoisier’s experiments [3]. The next generation of scientists, such as Carl and Gerty Cori and others proved that lactate and other metabolites produced by glucose metabolism in muscle could be used as substrates to replenish glucose in liver. The cycle wherein glucose is used by peripheral tissue for glycolysis and the resultant lactate is cleared by liver to resynthesize glucose is known as the Cori cycle.

Hepatic glucose synthesis, storage, and release are under the elegant control of hormone signaling networks unimaginable to early investigators. The discovery of insulin as a critical pancreatic hormone regulating glucose metabolism, and its lack as the cause of type 1 diabetes ranks among the most important medical breakthroughs of the twentieth century. Banting, Macleod and colleagues were awarded the Nobel prize in 1922, only a year after they discovered that pancreatic insulin extracts prevented death in type 1 diabetics [4]. Counterregulatory hormones, including glucagon, epinephrine, and certain glucocorticoids, which counter the actions of insulin by stimulating hepatic glucose production were discovered soon after. A major advance in understanding the mechanism of hormone action came in 1948 when Sutherland and coworkers demonstrated the existence of a glucagon-sensitive “glycogenolytic cascade” of kinase reactions which activates glycogen phosphorylase and controls the release of glucose from liver [5]. In a span of 150 years, vitalism was replaced by metabolism and the liver was identified as the source of a life-sustaining carbohydrate.

During the last century our understanding of hepatic carbohydrate metabolism grew exponentially, and broadened to include roles in lipid metabolism. The exact biochemistry of gluconeogenesis, the process that converts simple substrates into glucose, was defined. The complexities by which the gluconeogenesis is regulated are still being sorted out by researchers today. In addition, the duality of liver metabolism, embodied by its ability to both make glucose and oxidize lipid during fasting, but consume glucose and synthesize lipid upon feeding became a unique and defining feature of liver physiology. This fed-to-fasted transition is marshaled by the actions of insulin, glucagon, and other counterregulatory hormones. At the molecular level, the structure and function of enzymes involved in glucose metabolism

---

1 This chapter is an update of the same chapter in the 3rd edition of the International Textbook of Diabetes. The author is grateful to Christopher B. Newgard for his generous permission to retain certain excerpts of text from the previous edition [1].

Chapter 13

Lactate/Carbohydrate metabolism in the liver regulates blood glucose regardless of nutritional state or depletion. As fasting is extended, glycogen stores are depleted and gluconeogenesis becomes a greater contributor to glucose production. After an overnight fast in humans, glycogenolysis and gluconeogenesis contribute about equally to hepatic glucose production following an overnight fast [7]. After 48 hours of fasting, glycogen is depleted and glucose is produced almost exclusively by gluconeogenesis [7].

Many diseases disrupt hepatic glucose metabolism, but none impact modern society more than diabetes. Caused by insulin insufficiency (type 1 diabetes) or insulin resistance (type 2 diabetes), elevated blood glucose is a hallmark of these diseases. The lack or ineffectiveness of insulin and the prevailing actions of counterregulatory hormones like glucagon cause the simultaneous overproduction and impaired disposal of glucose by the liver (Figure 13.1). Thus, following sugar ingestion, poorly controlled type 1 diabetics are unable to suppress hepatic glucose production, have impaired glucose uptake and rapidly develop hyperglycemia. In addition the total loss of insulin and ablation of glucose catabolism potently activates hepatic fat catabolism and induces excessive ketone production and ketoacidosis that can be fatal in untreated type 1 diabetes. In type 2 diabetes, insulin levels and action are partially retained but defects of hepatic glucose production and uptake persist. Despite impaired insulin signaling to pathways which suppress glucose production, tangential signaling pathways continue to stimulate the conversion of glucose to lipid and cholesterol—a so-called “insulin paradox.” This residual insulin action and other emerging mechanisms promote dyslipidemia as a significant feature of insulin resistance. Thus, both forms of diabetes cause elevated hepatic glucose production and impaired glucose uptake, but impinge differently on the way hepatic glucose and lipid metabolism interact.

The important role of dysregulated hepatic glucose balance in the pathology of diabetes has motivated more than a century of research into the fundamental regulatory mechanisms of liver glucose metabolism. Over the past decade, modern tools of molecular biology, genetics, biochemistry, physiology, and metabolomics/flux have provided a vast amount of insight into regulatory mechanisms of hepatic glucose metabolism. This chapter is an extension of the third edition of the International Textbook of Diabetes Mellitus [1] and devoted to integrating this new information with classic knowledge of regulatory paradigms of glucose metabolism in liver.

**Hepatic glucose production**

**Overview**

Typical lean humans spend more than half of their lives in the post-absorptive state, with less than 5 g of glucose circulating in their blood to support life [6]. Many tissues rely on glucose as their primary fuel source. Notable examples are brain, which has limited access to fatty acids, and erythrocytes which do not possess mitochondria and, therefore, rely on glycolysis to meet energy requirements. Even during rest the body uses roughly 8 g of glucose per hour, and during exercise this rate can increase more than twofold. The body would deplete circulating glucose in less than 30 min, resulting in severe hypoglycemia, loss of neurologic function and death, if not for a constant endogenous supply of glucose. Under most conditions, liver provides approximately 90–95% of circulating glucose. This hepatic glucose production is maintained at precisely the same rate as whole body glucose utilization, keeping blood glucose between 80–100 mg dL⁻¹ regardless of nutritional state or activity level.

Hepatic glucose production is supplied by the breakdown of stored glycogen (glycogenolysis) and the synthesis of new glucose from noncarbohydrate precursors (gluconeogenesis). In humans, glycogenolysis and gluconeogenesis contribute equally to hepatic glucose production following an overnight fast [7]. After 48 hours of fasting, glycogen is depleted and glucose is produced almost exclusively by gluconeogenesis [7].
Inasmuch as most mechanistic studies referenced in this chapter have been carried out in rodent models, it is important to note that rats and mice deplete hepatic glycogen much faster, over approximately 24 and 12 h, respectively. Gluconeogenesis and glycogenolysis are regulated by hormone action and autonomic mechanisms that alter substrate supply, allosterie, posttranslational modification and enzyme transcription. These regulatory mechanisms are disrupted during type 1 and 2 diabetes leading to an inability to suppress gluconeogenesis and/or glycogenolysis and the development of hyperglycemia.

The metabolic mechanism of hyperglycemia during diabetes is one of the most widely studied pathologic features of any disease. Hyperglycemia (>124 mg/dL glucose) occurs when the equilibrium between hepatic glucose production and peripheral glucose utilization is disrupted. Either impaired glucose utilization or elevated hepatic glucose production can cause hyperglycemia (Figure 13.1). Many but not all studies find endogenous glucose production to be increased during type 2 diabetes. Perhaps more pertinent for type 2 diabetes is the failure of hyperinsulinemia to suppress hepatic glucose production, which is indicative of hepatic insulin resistance, particularly with regard to glucose metabolism [8]. Hepatic insulin resistance can be quantified using hyperinsulinemic-euglycemic clamp approaches [8]. Exogenous insulin is administered to achieve hyperinsulinemia while glucose is simultaneously infused to maintain normoglycemia. The rate of glucose infusion required to maintain normoglycemia reflects both insulin-mediated glucose disposal and suppression of hepatic glucose production. If an isotope tracer of glucose is co-infused, the rate of glucose appearance can be determined from the dilution of the tracer [8]. Type 2 diabetic humans have an impaired ability to suppress both hepatic gluconeogenesis and glycogenolysis [9] in response to insulin, and this revelation has led to extensive efforts to understand the metabolic, hormonal, and transcriptional regulation of these pathways using genetically malleable models such as mice.

**Glycogenolysis**

Hepatocytes store glucose as polymeric units called glycogen. Fully formed glycogen particles contain thousands of glucose molecules, and can have a molecular mass in excess of $1 \times 10^{6}$ Da. Glycogen synthesis is an important mechanism of hepatic glucose disposal during feeding, and will be discussed later. Glycogen degradation, or glycogenolysis, converts stored glycogen into glucose during fasting. This process is initiated by removal of glucose residues one at a time from the outer, nonreducing termini of the glycogen particle. Glycogenolysis is catalyzed by the active form of glycogen phosphorylase (phosphorylase a). This enzyme catalyzes phosphorolytic cleavage of the α-1,4-glycosidic bonds of the glucose polymer to yield glucose-1-phosphate. Glucose-1-phosphate is then converted to glucose-6-phosphate by phosphoglucomutase. The G6P that is formed from glycogenolysis is hydrolyzed to free glucose via the G6Pase complex, and then transported into the circulation by GLUT2 and other emergent mechanisms. The glycogenolytic cascade is activated in liver in the fasting state by a falling insulin concentration and rising glucagon concentration. The former reduces glycogen synthesis and the latter induces glycogenolysis (Figure 13.2). However, the switch between the two conditions may not be instantaneous or complete, so glycogen synthesis and degradation can occur simultaneously. This phenomenon is called “glycogen cycling” and is most often observed in the postprandial state or hyperglycemic state [10]. Glycogenolysis is essential for maintaining glucose supply to the brain and the central nervous system during early stages of fasting, ostensibly as a buffer until counterregulatory mechanisms fully activate glucose.

![Figure 13.2](image-url) Glycogen metabolism is regulated by the reciprocal actions of glycogen synthase and glycogen phosphorylase. Glycogenolysis is catalyzed by glycogen phosphorylase, which is positively regulated by glucagon-mediated phosphorylation and negatively regulated by allosteric interaction with glucose-6-phosphate. Glycogen synthesis is catalyzed by glycogen synthase, which is positively regulated by insulin and allosteric interaction with glucose-6-phosphate. The competing activity of these opposing reactions determines whether glycogen is synthesized or degraded to glucose.
gluconeogenesis. The human liver contains about 100 g of glucose as glycogen, which it depletes over about 48 hours of continuous fasting [11].

**Regulation of glycogenolysis**

Glycogen metabolism is regulated by reciprocal changes in the activities of glycogen synthase and glycogen phosphorylase (Figure 13.2). Net glucose production from glycogen occurs when phosphorylase activity is higher than synthase activity. Activation of glycogen phosphorylase is the culmination of a "glycogenolytic cascade" that is initiated by the binding of glucagon or β-adrenergic agonists to their receptors in liver cells [5]. The release of these hormones is stimulated by stressors such as fasting and exercise that require increased gluconeogenesis to maintain glycemia. Upon activation of their hepatocyte G-protein coupled receptors, and coordinated activation of adenylyl cyclase by the Gs heterotrimeric GTP-binding protein, glucagon causes a rapid rise in intracellular cAMP concentrations. This molecule causes the activation of cAMP-dependent protein kinase (PKA), which in turn leads to phosphorylation and activation of phosphorylase kinase. Finally, phosphorylase kinase phosphorylates glycogen phosphorylase, thereby converting the enzyme from its inactive (phosphorylase b) to its active (phosphorylase a) form (Figure 13.2) [12].

Similar to glycogen phosphorylase, glycogen synthase is phosphorylated under fasting conditions. However, unlike phosphorylase, phosphorylation of glycogen synthase occurs at multiple serine sites and inactivates rather than activates the enzyme [12]. Phosphorylation of glycogen synthase is catalyzed primarily by PKA and glycogen synthase kinase 3 (GSK3). In sum, a net activation of glycogenolysis is promoted during fasted or insulinopenic conditions via phosphorylation of glycogen phosphorylase and glycogen synthase, leading to activation of the former and decreased activity of the latter (Figure 13.2).

While glycogenolysis is promoted by phosphorylation of glycogen synthase and glycogen phosphorylase, a critical regulatory event for stimulating glycogen synthesis during feeding is to dephosphorylate these enzymes. The reciprocal regulation of glycogen synthase with phosphorylase and its overall participation in regulating glycogenolysis requires some consideration in our discussion of hepatic glucose production, although a more complete discussion of its role in glycogen synthesis and glucose disposal will be discussed later. A common protein phosphatase, protein phosphatase 1 (PP-1), is used to dephosphorylate both glycogen synthase and glycogen phosphorylase. The dephosphorylation mechanism is complemented by several other regulatory events. G6P binds to glycogen synthase, causing a conformational change that renders it a better substrate for PP-1. Similarly, glucose binds to glycogen phosphorylase a, again causing a conformational change that enhances its rate of PP-1-mediated dephosphorylation and decreases enzyme activity [12]. The expression of the glucose transporter GLUT2 and the glucose phosphorylating enzyme glucokinase in liver allows the levels of free glucose and G6P to rise in proportion to changes in circulating glucose concentrations, contributing to effective regulation of glycogen phosphorylase and glycogen synthase activities. In addition, the rise in insulin levels in the postprandial state activates a branch of the insulin signaling pathway that includes PI-3 kinase and Akt (protein kinase B) and stimulates phosphorylation and inhibition of glycogen synthase kinase 3 (GSK-3) [12], and other poorly understood mechanisms [13]. This results in less GSK-3-mediated phosphorylation of glycogen synthase, leaving the enzyme in a less phosphorylated, and therefore more active condition. The ability of glycogenolytic flux to be altered rapidly is an essential first response to ingestion of carbohydrate or upregulation of gluconeogenesis.

The common use of PP-1 for dephosphorylation of both glycogen synthase and glycogen phosphorylase is not a coincidence, and is an example of spatial organization in the regulation of hepatic carbohydrate metabolism. It is now appreciated that enzymes of glycogen metabolism, including the regulatory enzyme PP-1, are assembled in a multiprotein complex that is also associated with the glycogen particle. This spatial association also contributes to appropriate activation or suppression of glycogen synthesis in response to changes in nutritional conditions. The scaffolding proteins that organize glycogen metabolism are known collectively as glycogen-targeting subunits of PP-1 [12].

In principle, glycogen synthesis and glycogenolysis are reciprocally regulated by the antithetic control of glycogen synthase and glycogen phosphorylase (as discussed earlier). It should be noted, however, that this reciprocal regulation may not always be complete. Thus, in controlled experimental circumstances, such as during hyperinsulinemic-euglycemic clamp protocols in humans, a decrease in net hepatic glycogenolysis can be demonstrated, which is due to a large activation of glycogen synthase flux, but without a decrease in phosphorylase flux [14]. Thus, glycogenolysis can continue during periods of net glycogen synthesis, suggesting that a significant portion of the phosphorylated glucose derived from glycogenolysis in liver can be recycled back into glycogen (i.e., glycogen cycling). Thus, control of liver glycogen metabolism should not be viewed simply as regulation of perfectly reciprocal "on" and "off" switches, but more as control of net flux via a predominance of activity of one pathway relative to the other.

**Gluconeogenesis**

**Gluconeogenic pathways**

Gluconeogenesis is the synthesis of glucose from three carbon precursors such as alanine, pyruvate, lactate, and glycerol and is essentially a reversal of glycolysis [15] (Figure 13.3). Under normal conditions, 90% of gluconeogenesis occurs in liver, and the rest occurs mainly in the renal cortex. Other sites of gluconeogenesis, such as the small intestine have been suggested but remain controversial. Estimates of gluconeogenesis have...
Regulation of glucose metabolism in liver

Glucose metabolism is a complex process involving numerous enzymes and pathways. Key enzymes include glucokinase (GK), glucose-6-phosphatase, phosphofructokinase (PFK), fructose-1,6-bisphosphatase (FBP), pyruvate kinase (PK), and pyruvate carboxylase (PC). These enzymes catalyze the reversible reactions that convert glucose into glycogen and vice versa.

**Gluconeogenesis** and glycogenolysis are opposing pathways but share a majority of their enzymes. However, several steps are functionally irreversible and require separate enzymes. Glucokinase (GK) and glucose-6-phosphatase catalyze the opposing phosphorylation and de-phosphorylation of glucose. Phosphofructokinase (PFK) and fructose-1,6-bisphosphatase (FBP) catalyze the phosphorylation and de-phosphorylation of fructose-6-phosphate. The conversion of phosphoenolpyruvate (PEP) to pyruvate is catalyzed by L-pyruvate kinase (PK), but requires additional steps to synthesize PEP from pyruvate. Pyruvate must be transported into the mitochondria by the mitochondrial pyruvate carrier (MPC), then converted to oxaloacetate by pyruvate carboxylase (PC), and transported back to the cytosol where it is converted to PEP by phosphoenolpyruvate carboxykinase (PEPCK).

The opposing enzyme steps are important points of transcriptional and posttranscriptional control of gluconeogenesis.

Gluconeogenic enzymes

A minimum of 11 enzymatic steps are required to convert two molecules of pyruvate to one molecule of glucose [15]. Seven of these enzymes catalyze reversible reactions of glycolysis such as aldolase and triose phosphate isomerase. Another four reactions are catalyzed by unique enzymes which circumvent irreversible steps of glycolysis (Figure 13.3). Gluconeogenesis from pyruvate begins with the transport of pyruvate into the mitochondria via the mitochondrial pyruvate carrier (MPC). Once in the mitochondria pyruvate is converted to oxaloacetate, by the mitochondrial enzyme pyruvate carboxylase (PC).

**Gluconeogenic substrates**

Most amino acids, organic acids and glycerol can be substrates for hepatic gluconeogenesis. Notably, liver lacks the enzymes to convert acetyl-CoA to glucose, thus even chain fatty acids, ketones and certain ketogenic amino acids cannot be used as gluconeogenic substrates, although mixing of carbon pools in the TCA cycle does occur [16] (Figure 13.3). Lactate and alanine are key components of the Cori cycle, the process by which glycolysis in peripheral tissue produces these intermediates and the circulation delivers them to the liver where they are converted to pyruvate and used to resynthesize glucose. Glycerol is produced by lipolysis in adipose tissue during fasting, and in humans contributes to about 10% of glucose production after an overnight fast, though this number varies depending on the method used [16]. During continued fasting, endogenous glucose production remains constant because gluconeogenesis rises in proportion to compensate for glycogen depletion. After about 48 hours of fasting in humans, all glucose is produced by gluconeogenesis. This is roughly the same timing at which people enter ketosis, a metabolic adaptation of liver to produce ketones from lipids and supplement glucose utilization in peripheral tissue. During extended fasting, total glucose production decreases as low insulin levels suppress peripheral glucose utilization, hepatic glycogen is depleted and the body adapts to ketone and lipid oxidation.

**Figure 13.3** Gluconeogenesis and glycogenolysis are opposing pathways but share a majority of their enzymes. However, several steps are functionally irreversible and require separate enzymes. Glucokinase (GK) and glucose-6-phosphatase catalyze the opposing phosphorylation and de-phosphorylation of glucose. Phosphofructokinase (PFK) and fructose-1,6-bisphosphatase (FBP) catalyze the phosphorylation and de-phosphorylation of fructose-6-phosphate. The conversion of phosphoenolpyruvate (PEP) to pyruvate is catalyzed by L-pyruvate kinase (PK), but requires additional steps to synthesize PEP from pyruvate. Pyruvate must be transported into the mitochondria by the mitochondrial pyruvate carrier (MPC), then converted to oxaloacetate by pyruvate carboxylase (PC), and transported back to the cytosol where it is converted to PEP by phosphoenolpyruvate carboxykinase (PEPCK). The opposing enzyme steps are important points of transcriptional and posttranscriptional control of gluconeogenesis.

been based on either arterio-venous differences across the splanchnic bed, magnetic resonance measurements of hepatic glycogen depletion during fasting, or a variety of isotope tracer methodologies. In humans, gluconeogenesis contributes 40–50% of endogenous glucose production after an overnight fast, though this number varies depending on the method used [16]. During continued fasting, endogenous glucose production remains constant because gluconeogenesis rises in proportion to compensate for glycogen depletion. After about 48 hours of fasting in humans, all glucose is produced by gluconeogenesis. This is roughly the same timing at which people enter ketosis, a metabolic adaptation of liver to produce ketones from lipids and supplement glucose utilization in peripheral tissue. During extended fasting, total glucose production decreases as low insulin levels suppress peripheral glucose utilization, hepatic glycogen is depleted and the body adapts to ketone and lipid oxidation.
Oxaloacetate is then decarboxylated by phosphoenolpyruvate carboxykinase (PEPCK) to yield the glycolytic intermediate phosphoenolpyruvate (PEP). There are two isoforms of PEPCK, a cytosolic isofom (PEPCK-C) and a mitochondrial isofom (PEPCK-M). The human liver expresses about 50% of each, but the mouse liver expresses roughly 95% of its PEPCK as the cytosolic isofom [19]. Studies in genetically engineered mice indicate that the cytosolic isofom is most important for gluconeogenesis [20].

Since mitochondria do not possess an oxaloacetate transporter, a modified malate-aspartate shuttle is utilized to transport substrates from mitochondria to the cytosol where PEPCK-C and other gluconeogenic enzymes are located. The exact form of this shuttle depends on cytosolic redox state. In its simplest form, mitochondrial oxaloacetate is reduced to malate by mitochondrial malate dehydrogenase with the oxidation NADH to NAD+. Malate is transported out of the mitochondria in exchange for inorganic phosphate and then oxidized by cytosolic malate dehydrogenase to generate oxaloacetate and an NADH. Thus this pathway predominates when oxidized gluconeogenic substrates like pyruvate or alanine are utilized. Alternatively, mitochondrial oxaloacetate can be transaminated to aspartate and transported to the cytosol in exchange for glutamate, and then transaminated back to oxaloacetate. Although this process requires additional steps for transamination, it allows oxaloacetate to be transported in a redox neutral fashion (i.e., no net transport of NADH to the cytosol). The transaminase-dependent shuttle is critical for reduced substrates like lactate, but is not required for pyruvate itself.

Cytosolic phosphoenolpyruvate is converted to fructose-1,6-bisphosphate via six enzymatic steps common to glycolysis. However, conversion of fructose-1,6-bisphosphate to fructose-6-phosphate requires a distinct enzyme, fructose-1,6-bisphosphatase, as the ATP-consuming phosphofructokinase reaction of glycolysis is not reversible. Fructose-6-phosphate is then converted to glucose-6-phosphate by reversal of the hexose phosphate isomerase reaction of glycolysis. The terminal step of gluconeogenesis is the hydrolysis of glucose-6-phosphate (G6P) to free glucose, catalyzed by the glucose-6-phosphatase (G6Pase) enzyme complex. The complex is comprised of a catalytic subunit sequestered within the endoplasmic reticulum (ER), a glucose-6-phosphate translocase known as T1 that delivers glucose-6-phosphate to the catalytic subunit, and putative ER glucose and inorganic phosphate transporters (T2, T3) that move the reaction products back into the cytosol. The sequestration of G6Pase in the ER may protect other phosphorylated sugars from the nonspecific phosphohydrolase activity of its catalytic subunit. Specificity of the system is conferred by the T1 translocase component, which transports glucose-6-phosphate, but not its closely related epimer mannose-6-phosphate. This prevents uncontrolled hydrolysis of phosphorylated sugars and spares energy.

**Transcriptional regulation of gluconeogenesis**

The gradual shift from the fed-to-fasted state is followed closely by hormonal regulation of genes encoding PEPCK, fructose-1,6-bisphosphatase, and the glucose-6-phosphatase catalytic subunit. The transcriptional regulation of these genes remains a developing research area and its many emerging nuances cannot be completely covered here. However, this regulation is accomplished largely through two transcription factors, CREB and FOXO, the nuclear receptors GR, HNF-4α, and the coactivator PGC-1α [21,22] (Figure 13.4). Classic glucoregulatory hormones such as insulin, glucagon, and glucocorticoids mediate gluconeogenic gene expression by promoting or inhibiting the interaction of these proteins with the promoter regions of these genes [23]. More recently described hormones such as adiponectin, and posttranslational mediators of hepatocyte energy status like AMPK [24] and surtus also act on these transcriptional regulators to influence glucose metabolism in liver, and are discussed later.

**Induction of gluconeogenic gene expression**

Fasting causes increased levels and stability of PEPCK mRNA [25,26]. The PEPCK gene promoter contains binding sites for more than a dozen transcription factors, nuclear receptors and co-regulators. This allows the promoter to be exquisitely responsive to nutrition and stress. The best-studied factors regulating PEPCK expression are its activation by glucagon and glucocorticoids, and its inhibition by insulin [25,26] (Figure 13.4). Glucagon activates its G-protein coupled receptor on the cell surface of hepatocytes and stimulates cyclic AMP (cAMP) production. cAMP binds protein kinase A (PKA) and activates its kinase domain. After diffusion into the nucleus, PKA phosphorylates Ser133 of the cAMP-response element binding protein (CREB). CREB then binds to the cyclic AMP response element (CRE) in the PEPCK and G6Pase gene promoters [27]. Binding of phosphorylated CREB to its CRE site nucleates the assembly of its co-activator, the CREB binding protein (CBP), and initiates transcription. The stability of this complex is modulated through phosphorylation of the interacting domains by several kinases, including Ca2+/calmodulin-dependent kinases [21].

Transcription by CREB is also robustly potentiated by a second co-activator, cAMP-regulated transcriptional co-activator (CRTC2, or TORC2) (Figure 13.4). CRTC2 is maintained inactive by salt inducible kinases (SIK) which phosphorylate and sequester CRTC2 in the cytoplasm [21]. In addition to activating CREB, PAK also inhibits SIK, while other phosphatases, including calcineurin (a Ca2+/calmodulin-dependent phosphatase) dephosphorylate CRTC2. Dephosphorylated CRTC2 is then translocated into the nucleus and promotes transcription of CREB targets [21]. In addition to gluconeogenic genes, CREB also initiates the expression of other positive regulators of gluconeogenesis such as PGC-1α and NR4A, a nuclear receptor that also induces gluconeogenic gene expression. Thus, glucagon is a powerful impetus for regulating gluconeogenic
Regulation of glucose metabolism in liver

Figure 13.4 Transcriptional regulation of gluconeogenesis. Insulin and counterregulatory hormones instigate transcriptional events that modify the expression of key gluconeogenic enzymes. Glucagon mediates nuclear PKA activity which phosphorylates CREB, stabilizes the CRE complex and induces gluconeogenic enzyme expression. Insulin opposes gluconeogenic enzyme expression by signaling the phosphorylation of CRCT2 and FOXO1 which are then excluded from the nucleus and unable to induce gluconeogenic enzyme expression. Other factors such as glucocorticoids also have nuclear receptors that induce expression of gluconeogenic enzymes.

gene expression in liver, but its complexity also allows for its actions to be moderated or magnified by other factors, as we will see later. Employing a similar paradigm, glucocorticoids bind to the cytosolic glucocorticoid receptor (GR) causing its dissociation from a chaperone protein and its translocation to the nucleus where it binds to a glucocorticoid response element (GRE) in the PEPCK promoter (Figure 13.4). Formation of the GR/GRE complex leads to recruitment of other proteins (albeit different ones from those bound to the CREB/CRE complex), again leading to stimulation of PEPCK gene transcription.

Repression of gluconeogenic gene expression
Insulin represses glucagon and glucocorticoid induction of gluconeogenic gene expression through incompletely understood mechanisms that may be partially permissive in nature [28] (Figure 13.4). In contrast to glucagon action, which facilitates the binding of CREB to its promoters, insulin signaling excludes the forkhead transcription factor (FOXO1) from the nucleus and prohibits its binding to the promoters of PEPCK and G6pase [23]. Mouse studies in which insulin signaling and FOXO1 are acutely deleted have demonstrated that the upregulation of gluconeogenic genes during loss of insulin action requires FOXO1 [28,29]. Binding of insulin to its receptor initiates phosphorylation of the insulin receptor substrates and PI3Kinase. These events trigger a broader phosphorylation cascade which mediates the metabolic effect of insulin. Akt1/2, also known as protein kinase B, is activated by PI3Kinase and is required for most of insulin’s actions on lipid and glucose metabolism [22]. Akt1/2 (mainly Akt2 in the normal mouse liver) acts on gluconeogenesis by phosphorylating FOXO1, which is then excluded from the nucleus, causing gluconeogenic genes to be repressed [22]. Notably, Akt1/2 also phosphorylate and deactivate proteins required for glucagon action such as SIK and PGC-1α (see later), and activates proteins required for lipogenesis [23].

Role of PGC-1α
PGC-1α is a co-activator that binds to multiple transcription factors and nuclear receptors during fasting and/or the energy challenged state [30]. Like other coactivators, PGC-1α acts on nuclear chromatin to facilitate the binding transcription factors to target genes. Although very low in liver of fed mice, PGC-1α increases dramatically in liver during fasting [31]. The full induction of gluconeogenesis requires functioning PGC-1α and several models of type 1 and type 2 diabetes have increased expression or activity of hepatic PGC-1α [32]. PGC-1α is activated in response to cAMP and CREB and is thus an important component of counterregulatory hormone signaling. Once expressed, PGC-1α activity is modified posttranslationally by phosphorylation and acetylation under the regulation of AMPK and sirtuins (see later). Active PGC-1α induces the gluconeogenic enzymes, PEPCK, fructose-1,6-bisphosphatase, and the catalytic subunit of the G6Pase complex by co-activating hepatocyte nuclear factor 4-α and FOXO1 [31] (Figure 13.4).
PGC-1α also mediates fat oxidation in liver during fasting by co-activating PPARs. The induction of energy production by PGC-1α is essential to support the energetic expense of gluconeogenesis during fasting and is an additional route of gluconeogenic control by PGC-1α [32].

In summary, the regulation of gluconeogenic gene expression is carried out by a rise in cAMP levels caused by counterregulatory hormones during fasting or exercise. These hormones activate CREB, increase PGC-1α expression and stimulate expression of gluconeogenic genes via FOXO1 and HNF4α. Their effects are antagonized by insulin in the fed state via Akt1-mediated phosphorylation of FOXO1, resulting in suppression of gluconeogenic gene expression.

**Acute regulation of gluconeogenesis**

Prior to discovery of the transcriptional mechanisms described earlier, the control of hepatic glucose metabolism was examined in great detail on the basis of substrate, allosteric, and posttranslational modification. These factors alter gluconeogenic flux rapidly (seconds to minutes) and are a critical first response to increased glucose demand. Much of the acute regulation of gluconeogenesis is based on the antithetic regulation of several gluconeogenic/glycolytic enzyme pairs that catalyze opposing pathways. The metabolic intersections of PC/PDH, PEPC/PK, G6Pase/glucokinase and fructose-1,6-bisphosphatase/phosphofructokinase are critical sites for the regulation of gluconeogenesis. Logically, alterations that inhibit the regulatory glycolytic enzymes (PDH, GK, PFK, or PK) will promote gluconeogenesis, while alterations that inhibit regulatory gluconeogenic enzymes (PC PEPC, G6Pase, or fructose-1,6-bisphosphatase will promote glycolysis.

**Allosteric and covalent modification**

Allosteric modification of enzyme activity occurs when an enzyme binds to a molecule, causing either an increase or decrease in the enzyme’s activity. Allosteric binding sites are separate from the catalytic domains and interact with their effector molecules by noncovalent bonding. Typically, electrostatic or hydrogen bonding between the effector molecule and amino acid residues of the allosteric binding pocket cause a change in the quaternary structure of the protein or otherwise change the binding constant of the catalytic site of the enzyme. Importantly, the effect on conformation and activity is essentially instantaneous and independent of the transcriptional mechanisms described earlier. In contrast, covalent modification occurs when new functional groups are added to amino acids of the enzyme. The effect of covalent modification also changes the conformation and therefore the activity of the enzyme. The most common types of covalent modifications of metabolic enzymes are phosphorylation and acetylation, though many others can also occur. Covalent modification requires the activity of other proteins such as kinases and/or phosphorylases to add or cleave covalently bound functional groups.

**Pyruvate dehydrogenase**

Mitochondrial pyruvate can either be decarboxylated by PDH to yield acetyl-CoA (lipogenesis/oxidation) or carboxylated by PC to yield oxaloacetate (gluconeogenesis) (Figure 13.5). These divergent fates of pyruvate are influenced by nutritional state and are reciprocally regulated by allosteric and/or covalent modification. PDH is part of a very large complex of proteins known as the pyruvate dehydrogenase complex (PDC). Inasmuch as glucose production is required for survival, hepatic PDH is normally kept in an inactive state by allosteric inhibition and phosphorylation. PDH activity is allosterically inhibited by acetyl-CoA, NADH, and ATP, factors that are replete in the fasted liver. PDH is also inactivated by a family of pyruvate dehydrogenase kinases (PDK) that phosphorylate PDH [33]. PDK-4 expression is decreased by insulin action, increased by counterregulatory hormones and its activity is stimulated by

![Figure 13.5 Regulation of mitochondrial pyruvate flux.](image)

Mitochondrial pyruvate can be converted to acetyl-CoA by pyruvate dehydrogenase (PDH) or oxaloacetate by pyruvate carboxylase (PC). Both enzymes are metabolically regulated by cellular energy status. Acetyl-CoA, ATP, and NADH signal an energy rich hepatocyte and allosterically activate pyruvate dehydrogenase kinase-4 (PDK4), which deactivates PDH by phosphorylation. Acetyl-CoA on the other hand is a powerful allosteric activator of pyruvate carboxylase. Thus high energy states facilitate gluconeogenesis and suppress glucose oxidation and/or lipogenesis. In the overnight fasted state, PC is 10-fold more active than PDH.
acetyl-CoA, NADH, and ATP [33]. PDH activity is de-repressed during feeding, allowing glucose to be converted to acetyl-CoA for oxidation or lipogenesis (see hepatic glucose disposal later).

**Pyruvate carboxylase**

Suppression of PDH during starvation preserves pyruvate for pyruvate carboxylase (PC) and gluconeogenesis (Figure 13.5). PC is a mitochondrial enzyme that catalyzes the carboxylation of pyruvate to oxaloacetate and exerts more powerful control over gluconeogenesis than all of the other enzymes of the pathway combined [34]. Although its transcriptional regulation is weak, the allosteric control of PC has been thoroughly examined and provides a glimpse into the elegant regulation possible by acute metabolic feedback. In contrast to PDH, PC is allosterically activated by acetyl-CoA [35]. Thus, during fasting, when fat oxidation is increased, PC is activated and provides ample oxaloacetate for gluconeogenesis. Isotope tracer studies indicate that the *in vivo* activity of PC in the human postabsorptive liver is roughly up to 100 times higher than PDH [36,37].

The rapid regulation of PC activity by allosterism is made possible by its tetrameric structure [35]. PC has two biotin domains on each face of a tetrahedral. One domain facilitates the MgATP/HCO$_3^-$ dependent carboxylation of a biotin, and the other is a transferase domain where the carboxyl group is transferred to pyruvate. Activity of the PC complex requires that the carboxylated biotin domain access the transferase domain of the adjacent monomer. Acetyl-CoA binds amino acid residues in an allosteric binding site and induces a conformation that is favorable for the interaction of these domains. In the absence of acetyl-CoA, these domains cannot interact and PC activity approaches zero. The $K_a$ for acetyl-CoA binding is 20–60 uM, a value within the biologic range of acetyl-CoA concentration in liver mitochondria [35]. Intermediates associated with the TCA cycle such as α-ketoglutarate and glutamate reduce the $K_a$ for acetyl-CoA, thereby effectively inhibiting PC activity. This may provide a metabolic mechanism allowing oxaloacetate production to match TCA cycle capacity. Despite the elegant metabolic regulation of PC, its role as a rate-controlling step in gluconeogenesis has received little attention in the study of mechanism leading to the elevation of hepatic glucose production during diabetes.

**Pyruvate kinase**

Liver pyruvate kinase (L-PK) catalyzes the irreversible conversion of PEP to pyruvate and ADP to ATP. L-PK contributes to the positive regulation of glycolysis and lipogenesis and is discussed more thoroughly in these contexts in the later section on glucose disposal. However, it is worth noting that by opposing the combined actions of PC and PEPCK, L-PK acts as an important negative regulator of gluconeogenesis [15,38]. Experimentally, the generation of pyruvate from gluconeogenic PEP is an active substrate cycle in liver (Figure 13.6). Tracer experiments in hepatocytes [39], isolated liver [40], animal models [41] and humans [37] indicate that roughly half of PC flux is cycled back to pyruvate. Collectively these pathways are referred to as pyruvate cycling, and ostensibly confer greater metabolic flexibility by directing an already active flux towards or away from gluconeogenesis on fast timescales [15]. L-PK is suppressed by glucagon (and to a lesser extent epinephrine) via cAMP-mediated PKA action and as with other examples of actions of counterregulatory hormones, insulin opposes this process. L-PK activity is also allosterically activated by F-1,6-P2 and inhibited by alanine and ATP. Phosphorylation of L-PK by PKA decreases the maximal activation of L-PK by its substrates and potentiates the effectiveness of alanine and ATP to allosterically suppress its activity. Moreover, L-PK is a better substrate for phosphorylation by

---

**Figure 13.6** Metabolic regulation of liver pyruvate kinase. Pyruvate kinase opposes gluconeogenesis by removing PEP from the gluconeogenic pathway. Counterregulatory activation of cAMP and induction of PKA suppresses L-PK by rapid phosphorylation. Insulin promotes L-PK activity by opposing its phosphorylation and by inducing the glycolytic intermediate fructose-1,6-bisphosphate and allosteric activation of L-PK. The gluconeogenic precursor alanine has a negative allosteric effect on L-PK, consistent with a feed forward activation of gluconeogenesis. Tracer studies indicate that L-PK is constantly on and constitutes a pyruvate cycle between PC, PEPCK and L-PK, thereby conferring greater metabolic control of PEP entrance into gluconeogenesis.
PKA when it is allosterically inhibited by alanine or ATP than when it is allosterically activated by F-1,6-P2. Thus allosteric regulation not only mediates L-PK activity directly, but also deters phosphorylation of the enzyme [15].

**Fructose-1,6-bisphosphatase/ phosphofructokinase**

Allosteric regulation of hexose flux is another example where the regulation of gluconeogenesis occurs by interaction with glycolytic regulation. The conversion of F-1-P to F-1,6-P2 is irreversibly catalyzed by the gluconeogenic enzyme fructose-1,6-bisphosphatase, and the opposite glycolytic reaction is reversibly catalyzed by phosphofructokinase. Like the intersection between PC-PEPCK and L-PK, fructose-1,6-bisphosphatase and PFK resemble a futile cycle but which is actually a finely tuned point of metabolic control [15]. Liver phosphofructokinase is allosterically inhibited by PEP and citrate, which are indicative of abundant energy and substrate for gluconeogenesis (Figure 13.6). More importantly, phosphofructokinase is regulated by F-2,6-P2, a product of F-1,6-P2, and the bifunctional enzyme. Fructose-1,6-bisphosphatase is also inhibited by F-2,6-P2, the product of the bifunctional enzyme during feeding [15]. However, like most other substrate cycles discussed in this chapter, complete deactivation of the opposing fluxes of fructose-1,6-bisphosphatase and the bifunctional enzyme is not complete, even during a substrate load [42]. The regulation of the relative activities of FBPase and PFK is addressed again in the discussion of hepatic glucose disposal later.

**Hepatic energetics and glucose production**

Hepatic gluconeogenesis uses more than 40% of the energy consumed by the human liver [36]. Gluconeogenesis consumes 4 moles of ATP, 2 moles of GTP, and 2 moles of NADH for every mole of glucose produced. This energy is supplied in large part by the oxidation of fatty acids, which is promoted in the fasted state via a fall in the concentration of malonyl CoA and activation of carnitine palmitoyltransferase 1 (CPT 1) [43]. The allosteric and substrate effectors mentioned throughout this section (NADH/NAD+, ATP/AMP, acetyl-CoA, citrate, etc.) are products of fatty acid oxidation. Consequently, fat oxidation is required for optimal gluconeogenesis in hepatocytes and perfused liver [44]. When defects in fat oxidation occur, either due to targeted ablation in mice or inborn errors in humans, impaired gluconeogenesis and hypoglycemia almost always emerge as complications. In contrast, exogenous lipid infusion induces gluconeogenesis in vivo [45]. The metabolic dependence of hepatic gluconeogenesis on fat oxidation is imparted by the allosteric and covalent activation of the enzymes of gluconeogenesis and glycolysis [44] (Figure 13.7). Recall that PC activity is allosterically activated by acetyl-CoA, a product of the β-oxidation of fatty acids. Condensation of acetyl-CoA with oxaloacetate is the first step of the tricarboxylic acid (TCA) cycle and leads to the generation of citrate and NADH, and ultimately to a rise in ATP/ADP through the actions of the respiratory chain. Together, citrate, NADH and ATP, are the same factors that suppress PFK, PK, and PDH in glycolysis. In addition, the high energy cofactors, ATP, GTP, and NADH, are consumed by the gluconeogenic enzymes, PC, PEPCK, phosphoglycerate kinase, and glyceraldehyde-3-phosphate dehydrogenase (GAPDH), PEPCK, and phosphoglycerate kinase (PGK), respectively.

**Figure 13.7** Hepatic energy production positively enhances gluconeogenesis. The metabolic products and cofactors downstream of oxidative metabolism are metabolic activators of gluconeogenesis. Acetyl-CoA activates PC and inhibits PDH. Citrate generated during the oxidation of acetyl-CoA is a powerful inhibitor of the glycolytic enzyme PFK and therefore an activator of gluconeogenesis. NADH and GTP generated by the TCA cycle and ATP generated by respiration are required cofactors for the gluconeogenic reactions glyceraldehyde 3-phosphate dehydrogenase (GAPDH), PEPCK, and phosphoglycerate kinase (PGK), respectively.
by suppressing NADH utilization in the respiratory chain. Suppressed energy charge has secondary effects that reduce gluconeogenic gene expression through AMPK signaling [48], though this does not appear to be necessary for many of the metabolic effects of metformin [46]. A primary action on gluconeogenesis by metformin may be more metabolic in nature. By limiting energy supply, metformin inhibits high rates of gluconeogenesis, consistent with the requisite role of hepatic energy production in support of gluconeogenesis described earlier. In contrast to many other therapeutic targets of hepatic glucose metabolism, metformin has relatively few adverse effects, with rare manifestations of lactic acidosis being the most common.

**Substrate supply and hepatic glucose production**

The availability of gluconeogenic substrate is an important determinant of hepatic glucose production [49]. Glucagon and other hormones act on peripheral tissues to increase substrate delivery to liver. These events are particularly relevant during fasting and diabetes when a rise in the glucagon:insulin ratio causes activation of hormone-sensitive lipase and other lipases, which catalyzes the hydrolysis of triglycerides to yield free fatty acids and glycerol. The free fatty acids are transported to liver and re-esterified or oxidized. Oxidation is increased during fasting and diabetes because lower glycolytic flux in these states decreases malonyl-CoA mediated inhibition of CPT1 transport of fatty acids into mitochondria and increases the rate of fatty acid oxidation [43]. Fatty acid oxidation provides NADH and ATP required for optimal gluconeogenic activity [44]. Glucagon and glucocorticoid action also increases the supply of gluconeogenic precursors by inducing skeletal muscle autophagy, protein synthesis, and amino acid supply, while insulin has the opposite effect [50]. In addition counterregulatory hormones and fatty acids activate fat oxidation in muscle and suppress PDH activity which in turn stimulates aerobic glycolysis in the periphery. The lactate formed under these conditions is used by liver for gluconeogenesis and forms the basis of the Cori cycle. This cycle is particularly important during exercise where lactate formation in muscle provides ATP and defers the oxidative burden to liver.

The relative importance of substrate supply and hormone action on hepatic glucose production has been studied using mouse genetics and sophisticated metabolic clamps. Chronic ablation of the liver insulin receptor results in hyperglycemia [51], but acute ablation does not affect glycemia [52,53]. Specific restoration of the liver insulin receptor in mice missing the receptor in peripheral tissue is also insufficient to normalize hepatic glucose production [54]. Thus, the peripheral effects of insulin are important determinants for hepatic glucose metabolism. In canine models, where hormone and substrate can be delivered directly to liver by infusion into the portal vein, insulin action acutely suppresses hepatic glucose production by reducing glycolgenolysis and directing gluconeogenic G6P to glycogen synthesis [55]. Although hepatic insulin is sufficient to mediate changes in gluconeogenic gene expression, insulin per se, does not rapidly alter gluconeogenic flux [55]. Thus, insulin action in liver is effective at regulating glycogen flux in the course of minutes, but requires hours to establish the molecular conditions to increase hepatic gluconeogenesis. In contrast, experiments where insulin was delivered to peripheral tissue, gluconeogenesis was reduced mainly by insulin-mediated suppression of adipose lipolysis and reduced delivery of free fatty acids to liver [56]. Insulin’s effect on hepatic gluconeogenic gene expression and its ability to restrict substrate supply from the periphery work together to reduce gluconeogenesis.

**Linking molecular and metabolic regulation of hepatic glucose metabolism**

Cross-talk between metabolism and molecular biology is an emerging mechanism of metabolic regulation and is particularly relevant to the control glucose metabolism in liver. Some transcription factors and nuclear receptors are activated in response to metabolites. During fasting lipid metabolites are released that activate PPARα and induce the gene expression of oxidative metabolism. The activation of mitochondrial metabolism, as described earlier, has a positive effect on gluconeogenesis and ketogenesis [57]. In contrast, hypoxia inducible factor-1α (HIF-1α) is a transcription factor that upregulates glycolysis and suppresses mitochondrial respiration when TCA cycle intermediates (e.g. fumarate, succinate, and α-ketoglutarate) are elevated. Although most commonly related to mitochondrial dysfunction in heart disease and cancer, constitutive stabilization of hepatic HIF by loss of the von Hippel-Lindau protein results in dramatically impaired hepatic TCA cycle, respiration, blunted gluconeogenesis, hepatic steatosis, and hypoglycemic death [58]. The physiologic role of HIF in liver has not been rigorously studied. Other proteins, such as AMPK [59] and sirtuins [60], mediate the phosphorylation and acetylation of multitudes of enzymes, cell signaling proteins and transcription factors in response to cellular energy status.

**AMPK**

AMPK activated protein kinase (AMPK) is activated in response to elevated AMP (i.e., low energy). AMPK generally activates metabolic pathways that produce cellular energy and suppresses pathways that consume energy [24,59,61]. In liver, AMPK suppresses anabolic pathways like gluconeogenesis, lipogenesis, and cholesterol synthesis [61] and activates β-oxidation via PGC-1α [62]. Among AMPK’s most acute responses to energy demand is the rapid phosphorylation and inactivation of acetyl-CoA carboxylase (ACC) [63] resulting in lower malonyl-CoA, activated CPT1 and increased oxidative metabolism [64]. AMPK activation is exquisitely sensitive to the energetic stress of gluconeogenesis [65]. The normal induction of hepatic AMPK during fasting does not occur when gluconeogenesis is blunted [66]. Thus, among its many other functions, AMPK helps the liver match energy production to the energy demand of gluconeogenesis.
Sirtuins

Sirtuins are a class of NAD\(^+\) dependent deacetylase that regulate transcription and posttranslational activity by deacetylating key sites of numerous proteins [60]. There are 7 sirtuins in humans, with sirt1 and sirt3 playing important roles in the deacetylation of cytosolic/nuclear and mitochondrial proteins, respectively. Because of their dependence on NAD\(^+\), sirtuins are potentially redox sensitive and thus metabolically regulated. For example, the cytosolic/nuclear sirt1 upregulates gluconeogenesis by deacetylating several important regulatory factors including PGC1-\(\alpha\) and FOXO1. However, sirt1 also suppresses CREB-mediated gluconeogenic gene expression which may reduce gluconeogenesis during starvation as ketone production predominates and gluconeogenic precursors become limited. Mitochondrial sirt3 may influence gluconeogenesis by modifying the activity of \(\beta\)-oxidation and the TCA cycle [67].

Hepatic glucose disposal

Overview

Carbohydrate ingestion switches liver from a net glucose producer to a net glucose consumer [68]. This switch is rapid, and potentiated by specialized isoforms of glycolytic enzymes that promote glucose uptake and phosphorylation only when glucose concentration is high. In addition, feeding alters hormones that mediate the phosphorylation and expression patterns of glycolytic and gluconeogenic enzymes to favor catabolism of glucose. Net hepatic uptake of glucose (1) reduces circulating glucose concentration, (2) replenishes hepatic glycogen, and (3) converts carbohydrate to lipid for long-term energy storage (Figure 13.8). In this section we consider the factors that regulate hepatic glucose uptake and utilization.

Glucose uptake

A family of facilitated glucose transporters with unique tissue distribution and kinetic properties emerged from the work of several laboratories in the late 1980s and early 1990s. The major glucose transporter isoform of liver and pancreatic islets, GLUT2, was the second member of the family identified. All facilitated glucose transporters are equilibrium-based transporters that are capable of bidirectional glucose transport across cellular membranes, with the directionality determined by the relative intracellular and extracellular glucose concentrations. GLUT2 has a low affinity (\(K_m > 10 \text{ mM}\)) but a high capacity for glucose transport relative to other members of its gene family. Thus GLUT2 transports large amounts of glucose into the hepatocyte, but only when the glucose concentration is high (10 mM). In normal physiology, liver takes up significant amounts of glucose only during meal absorption.

Glucokinase

Glucose phosphorylation in liver is primarily catalyzed by glucokinase (hexokinase IV). Like GLUT2, glucokinase (GK) has a lower affinity for glucose and a higher catalytic capacity than the other members of its gene family (i.e., hexokinases I, II, and II) [69]. Glucokinase is expressed liver, the islets of Langerhans, and certain specialized neuroendocrine cells in the pituitary and gastrointestinal tract, whereas hexokinase I is found in brain, red blood cells and many other tissues, and hexokinase II is predominantly expressed in muscle and adipose tissue. Glucokinase has an \(S_0.5\) for glucose of about 8 mM, sigmoidal substrate dependency (indicated by a Hill coefficient of 1.7), and unlike hexokinases I and II, is not allosterically inhibited by the product of its reaction, glucose-6-P [69]. Thus, hepatocytes transport and phosphorylate glucose with kinetic features that allow net glucose uptake only during digestion.

In addition to the unique kinetics of hepatic glucokinase, its activity is also controlled by hormonal regulation of transcription and compartmental segregation. During fasting, glucokinase is sequestered in the nucleus by a glucokinase regulatory protein (GKRP). Ingestion of carbohydrate stimulates dissociation of glucokinase from GKRP and translocation of the enzyme from the nucleus to the cytosol where it can be active. Compartmentalization of glucokinase in the nucleus not only prevents glucose phosphorylation during gluconeogenesis, but also protects the enzyme from protein degradation. The acute induction of glucokinase by translocation to the cytosol is supplemented by a 20–30-fold increase in insulin-mediated glucokinase gene expression [70,71]. These large changes in glucokinase mRNA

Figure 13.8 Glucose uptake by the liver is modulated by insulin and glucose concentration. Specialized isoforms of glucose transporter (GLUT2) and hexokinase (GK) equip the liver to use a net amount of glucose only when glucose concentrations rise above normal, for example after a meal. The accompanying elevation in insulin stimulates glycolysis and the conversion of glucose into glycogen and lipids. Glycolytic intermediates (i.e., glucose-6-phosphate, fructose-1,6-bisphosphate, xyulose-5-phosphate, etc.) allosterically activate glycogen synthesis and lipogenesis. Glucose disposal by glycogen synthesis is enhanced by its indirect synthesis via glycolytic intermediates in liver or gluconeogenic intermediates imported from peripheral tissue.
are accompanied by more modest changes in glucokinase enzyme activity [71]. The time frame for changes in glucokinase enzyme activity due to insulin action is 30–60 min.

A key illustration of glucokinase in control of human fuel homeostasis comes from the discovery that a form of non-insulin-dependent diabetes mellitus (NIDDM) known as maturity-onset diabetes of the young (MODY) type 2 is caused by mutations in the glucokinase gene [72]. All MODY-2 patients are heterozygous, with one normal and one mutated glucokinase allele. Most, but not all, mutant enzymes associated with MODY-2 exhibit large decreases in enzymatic activity due to changes in kcat and Km for glucose and Km for ATP. MODY-2 patients have impaired hepatic glucose disposal, leading to an increase in net hepatic glucose production and reduced glycogen storage compared to normal subjects [73]. In addition, patients with MODY-2 have a defect in glucose-stimulated insulin secretion, due to the important role that glucokinase plays in regulating the rate of glycolysis and thereby the production of key stimulus/secretion coupling factors in islet β cells.

**Glucose phosphorylation as a therapeutic target for diabetes**

Early experiments using genetic overexpression of glucokinase in liver of rodent models showed promise for the protection against hyperglycemia during diabetes, spanning the development of glucokinase activators [74]. These drugs interact with allosteric sites of glucokinase, thus promoting a conformation that exposes the catalytic domain and potentiates the activity of the enzyme [74]. Glucokinase activators have achieved beneficial effects on glycemia in human trials, in part by the putative activation of hepatic glycogen storage (a pathway that is impaired in the diabetic state). However, chronic genetic gain of function also causes fatty liver, hyperlipidemia, and increased insulin resistance [75]. These undesired effects appear to be linked to the overstimulation of the lipogenic pathway by increased glucose uptake. Unfortunately, some of these long-term side effects have been recapitulated in clinical trials, leading to tempered enthusiasm for glucokinase activation [76]. A related approach is to reduce hepatic glucose output by inhibition of glucose-6-phosphatase. Pharmacologic inhibitors of glucose-6-phosphatase have not been widely tested, but mice with a liver-specific loss of function have a dyslipidemia phenotype similar to glucokinase activation, in addition to glycogen storage disease [77]. The negative effect of these interventions on lipidemia raises concerns that sequestering carbohydrate in liver may improve glycemia at the expense of promoting dyslipidemia.

**Glycogen synthesis**

The normal postabsorptive human liver stores roughly 100 g or approximately 300 mM of glucose as glycogen which it releases during fasting to maintain blood glucose concentration. By 48 h of fasting, the human will have depleted most of its liver glycogen. Following a meal, when hepatic glucose-6-phosphate is plentiful, glycogen can be replenished within several hours. The synthesis of glycogen begins with conversion of glucose-6-phosphate to glucose-1-phosphate by phosphoglucomutase, followed by the “activation” of glucose-1-phosphate to UDP-glucose by UDP-glucose pyrophosphorylase. Finally, glycogen synthase uses UDP-glucose to add glucose molecules one at a time to a growing glycogen particle. The mature glycogen particle has a highly branched structure, which is created by a branching enzyme that moves blocks of glucose residues and links them in β-1,6-glycosidic linkages. Approximately half of hepatic glycogen synthesized after a mixed meal originates by the “direct” conversion of glucose to glycogen. The remainder is synthesized by an “indirect” pathway, which refers to the observation that some exogenous glucose is metabolized to trioses (i.e., at least DHAP and GA3P but perhaps all the way to pyruvate) before being converted to glycogen. The indirect pathway of glycogen synthesis is the consequence of continued gluconeogenesis during the fasted-to-fed transition [78], and may be reinforced by Cori cycling. The indirect pathway can account for approximately 30–60% of liver glycogen synthesis in humans, with the exact percentage dependent on the time elapsed between meals [79].

Glycogen synthase exists as two major isoforms and is regulated by both allosteric and covalent modification [12]. GYS1 is expressed in muscle and other cells which store glycogen, while hepatocytes express mainly GYS2. Glycogen synthase was one of the first enzymes to be identified whose activity is modified by phosphorylation, which occurs on multiple serine residues by several different kinases, notably PKA and GSK3α. Glycogen synthase activity is generally suppressed by phosphorylation, especially at site 2 (ser 7) [12]. PKA phosphorylation (glycogen synthase deactivation) is triggered by glucagon receptor activation, while GSK3 activity is suppressed (glycogen synthase activation) by insulin. The latter is mediated by Akt phosphorylation of GSK3, though GSK3 independent mechanisms are also involved [13] and are likely mediated by allosteric regulation. Glycogen synthase is potently activated by allosteric interaction with glucose-6-phosphate [80]. Thus, conditions such as elevated portal glucose and insulin lead to increased glucose transport and glucokinase activity, which favor increased glucose-6-phosphate concentration and induction of glycogen synthesis (Figure 13.8).

Phosphorylated glycogen synthase (deactivated) requires regulatory phosphatases for reactivation. Phosphatase activity is organized by scaffolding proteins known as glycogen targeting subunits [81]. These proteins are part of a large family of more than 50 protein phosphatase-1 (PP1) binding proteins that deliver the enzyme to a wide array of substrates and cellular addresses, allowing PP-1 to participate in diverse cellular processes such as glycogen metabolism, cell division, vesicle fusion, and ion channel function. Gc is a 35-kDa glycogen targeting protein that is preferentially expressed in liver. Protein targeting to glycogen (PTG), also known as PPP1R5, and a fourth form,
PPP1R6, are similar in size to G\_l but are expressed in a wide variety of tissues. All of the targeting subunits are able to bind glycogen and PP-1, and exhibit varying capacities for binding of glycogen synthase, glycogen phosphorylase, and phosphorylase kinase [81]. Overexpression of PTG inhibits activation of glycogenolysis by the glycogenolytic cascade, and protects diabetic animals from hyperglycemia. However, constitutive activation of PTG also causes a form of glycogen storage disease in nondiabetic animals. PTG is particularly important for the regulation of indirect glycogen synthesis from gluconeogenic substrates [81]. Although the mechanism of PTG-mediated indirect synthesis of glycogen is not entirely clear, it appears to be related to accelerated disposal of glucose-6-P into glycogen, and a stimulation of gluconeogenesis. Thus PTG stimulates glycogen synthesis independent of glucose transport and GK activity.

**Glycogen metabolism as a therapeutic target for diabetes**

One approach to reducing hepatic glucose production during diabetes is to improve glycogen storage by activation of glycogen synthase and/or inhibition of glycogen phosphorylase. Glycogen synthase has been targeted by inhibition of GSK-3. GSK-3 inhibition increases glycogen storage and improves glycemia in a rodent model of type 2 diabetes [82]. However, glycogen synthase kinase also phosphorylates many other targets, making its mechanism difficult to decipher. Alternatively, glycogen phosphorylase can be inhibited directly by several classes of glucose analogues that either impair its allosteric activation or repress its ATP binding capability. These compounds improve glycemia in animal models of diabetes by increasing direct and indirect glycogen synthesis [83]. The adverse effects of pharmacologic promotion of glycogen storage are not completely known, but long term glycogen phosphorylase inhibition in animal models of type 2 diabetes can cause hepatomegaly, glycogenosis, and steatosis, side effects also common to glycogen storage diseases.

**Glycolysis**

Glucone-6-phosphate not used for direct glycogen synthesis can be metabolized in the glycolytic pathway. About half of the glucose taken up by liver during duodenal glucose absorption will undergo glycolysis to pyruvate. The resulting pyruvate can be used for lipogenesis and amino acid synthesis, lactate production, or cycled back to glucose-6-phosphate for indirect glycogen synthesis. There are 10 steps in the glycolytic pathway (Figure 13.3). Seven are catalyzed by enzymes with equilibrium constants that allow the reaction to proceed in either the “forward” (glycolytic) or “reverse” (gluconeogenic) direction dependent upon physiologic changes in the relative concentrations of substrates and products of the reactions. Three enzymatic steps, glucose phosphorylation by glucokinase (see earlier), the conversion of fructose-6-phosphate to fructose-1,6-bisphosphate by phosphofructokinase, and the conversion of phosphoenolpyruvate to pyruvate by pyruvate kinase are considered essentially irreversible due to a large release of free energy. Free energy is provided by the hydrolysis of ATP in the case of the glucokinase and phosphofructokinase reactions, and via the favorable thermodynamics inherent in the conversion of an enolphosphate to a ketone in the pyruvate kinase reaction. In the previous section of this chapter, we discussed distinct enzymes that have evolved to circumvent the otherwise irreversible glycolytic steps and allow gluconeogenesis to occur. Thus, the forward reactions of glycolysis serve to produce pyruvate and also moderate gluconeogenesis during feeding. However, despite the elegance of the regulatory mechanisms of glycolysis, liver is, first and foremost, a gluconeogenic organ and gluconeogenesis remains modestly active even during the unfavorable conditions of a glucose load [42]. This residual activity provides a mechanism to prevent lactate build-up, and is the basis for indirect glycogen synthesis.

**Regulation of phosphofructokinase (PFK)**

Glucone-6-phosphate is rapidly isomerized to fructose-6-phosphate, and then undergoes phosphorylation to fructose-1,6-bisphosphate by PFK with ATP serving as the phosphate donor. PFK displays sigmoidal kinetics, typical of enzymes that are regulated by allosteric ligands. Its activity is regulated by hepatic energy charge, such that increases in the ratio of ATP:ADP inhibits enzyme activity. Citrate, the product of the first committed step of the TCA cycle, is also a potent inhibitor of PFK activity. This interaction reduces glycolysis and increases gluconeogenesis when TCA cycle intermediates are replete and energy production is high. However, the most important allosteric regulator of PFK activity in liver is fructose-2,6-bisphosphate (F-2,6-P2) [84–86]. F-2,6-P2 activates PFK at low micromolar concentrations via two main actions. First, it converts the kinetic profile of PFK from sigmoidal to hyperbolic with respect to the concentration of the PFK substrate fructose-6-phosphate. This change in activity increases the affinity (lowers the Km) of PFK for fructose-6-phosphate. Second, F-2,6-P2 relieves inhibition of PFK by ATP, such that higher concentrations of the nucleotide are required to suppress enzyme activity.

Biochemical and molecular studies revealed that the synthesis and degradation of F-2,6-P2 is catalyzed by a single protein with two distinct catalytic sites [87]. This protein was named 6-phosphofructose-2-kinase/fructose-2,6-bisphosphatase but is also known more simply as “the bifunctional enzyme.” The bifunctional enzyme found in liver is regulated by fasting and glucagon action through cAMP-dependent PKA phosphorylation of a serine residue near its N-terminus [88]. Phosphorylation at this site inhibits the 2-kinase activity of the enzyme, and increases its bisphosphatase activity, resulting in a net decrease in F-2,6-P2 concentration within cells. This in turn leads to reduced affinity of PFK for its substrate fructose-6-phosphate and an increase in sensitivity of the enzyme to the inhibitory effects of ATP, leading to a net decrease in enzyme activity. Under these conditions, glycolysis is inhibited and gluconeogenesis is promoted.
In the transition from the fasted to the fed state, this regulation is reversed. Following a mixed meal, insulin and glucose levels rise in the blood while glucagon is decreased. These conditions cause dephosphorylation of the bifunctional enzyme, restoring its 2-kinase activity and causing a rapid increase in F-2,6-P2 levels. The dephosphorylation of the bifunctional enzyme appears to be mediated by a hexose-phosphate-sensitive protein phosphatase (PP2A). Xylose-5-phosphate, an intermediate of the pentose phosphate pathway, was first discovered to activate PP2A [89]. Thus, carbohydrate-regulated PP2A provides a direct link between the postprandial change in intracellular glucose concentration and the reactivation of the 2-kinase activity of the bifunctional enzyme. In addition to the effect of glucose to activate PP2A, the increase in insulin levels in the postprandial state serves to stimulate degradation of cAMP to AMP via phosphodiesterases, thus decreasing the activity of PKA and contributing to maintenance of the bifunctional enzyme in a nonphosphorylated state.

Hepatic glycolysis as a therapeutic target for diabetes
Genetic modification of the bifunctional enzyme, so that it favors formation of F-2,6-P2 and activation of PFK, induces hepatic glycolysis, reduces glucose production and improves glycemia in diabetic models, without the deleterious effects on circulating FFA and TG levels that occurs with glucokinase activation [90]. Although both glucokinase and PP2K activation stimulate hepatic glycolysis, unlike glucokinase activation, PFK activation appears to avoid induction of lipogenesis by preventing the accumulation of mono-phosphorylated glycolytic intermediates (i.e., xylose-5P and glucose-6P) that activate ChREBP [90]. However, studies have been limited to genetic activation in rodent models and as was also the case for glucokinase, a more detailed understanding of the metabolic consequences of activation of the bifunctional enzyme/PFK is required, including better understanding of changes in lipid metabolism and the potential for hypoglycemia.

Regulation of pyruvate kinase (PK)
PK catalyzes the conversion of PEP to pyruvate, the last of the irreversible steps of glycolysis in liver. As described in the discussion of its role in gluconeogenesis (see earlier), pyruvate kinase is subject to both allosteric and transcriptional regulation (Figure 13.6). For example, liver isoform of pyruvate kinase (L-PK) is activated by the product of the PFK reaction, fructose-1,6-bisphosphate. This represents a “feed-forward” mechanism by which flux through an early step in glycolysis activates a later step. Similar to PFK, L-PK is sensitive to the energy charge of the cell, and its activity is decreased in the presence of a high ATP:ADP + AMP ratio. Similar to the bifunctional enzyme, L-PK is inhibited by PKA-mediated serine phosphorylation [91]. Phosphorylation increases the apparent Km of the enzyme for its substrate PEP, and renders the enzyme more sensitive to inhibition by a high energy charge (elevated ATP:ADP + AMP ratio).

Like the other key glycolytic enzymes, transcription of the L-PK gene is decreased by fasting or insulinopenic diabetes, and is increased in the fed state or by insulin injection. Glucose appears to be a more direct and potent regulator of L-PK transcription than insulin, which may be acting to antagonize the effects of glucagon by activating phosphodiesterases and lowering cellular cAMP levels. In addition, the same protein phosphatase that dephosphorylates the bifunctional enzyme when sugar phosphate levels are elevated also dephosphorylates and activates ChREBP. ChREBP stimulates transcription of L-PK, as well as lipogenic genes such as acetyl-CoA carboxylase and fatty acid synthase. Thus, a rise in circulating glucose concentration and a proportional increase in the rate of hepatic glycolysis and pentose monophosphate shunt activity in the fasted-to-fed transition is translated into an increase in 2-kinase activity of the bifunctional enzyme via activation of PP2A, resulting in an increase in F-2,6-P2 levels and increased flux through PFK to generate fructose-1,6-bisphosphate. This intermediate then serves as a feed-forward activator of a later step in glycolysis, L-PK. Finally, allosteric activation of pyruvate kinase is supplemented by another action of the Xu5P-regulated protein phosphatase, the dephosphorylation and activation of the ChREBP transcription factor which stimulates L-PK gene transcription.

Pyruvate dehydrogenase and lipogenesis
The generation of pyruvate via glycolysis and L-PK brings the chapter full circle. Tracer experiments in humans estimate PDH flux to be very small in the postabsorptive liver [36,37]. However, in the fed state activation of hepatic lipogenesis requires activation of PDH flux to generate acetyl-CoA from carbohydrate. PDH consists of three subunits (E1–E3) that catalyze the sequential decarboxylation, acetyl-CoA formation and reduction of NAD+ to NADH, respectively. These subunits exist in a super-complex of more than 200 of these subunits called the PDH complex (PDC) [92]. During fasting, PDH activity is kept suppressed by pyruvate dehydrogenase kinase (PDK) which phosphorylates one or more of three serine residues on the E1 subunits [33]. PDK is associated to the E2 subunit of the PDC and its activity is acutely regulated by its own substrates and products (ATP/ADP) as well as the substrates and products of PDH (NADH/NAD+ and acetyl-CoA/CoA). Upon carbohydrate ingestion, PDH is reactivated by dephosphorylation of these sites through the actions of pyruvate dehydrogenase phosphatases (PDP), which are also components of the E2 subunits of the PDC and broadly regulated by Ca2+ and Mg2+. The main regulation of PDH phosphorylation appears to be through modification of PDK activity [33]. Upon carbohydrate ingestion, fatty acid oxidation is suppressed, PDK is deactivated, and PDPs dephosphorylate to relieve the suppression of PDH.

In coordination with the conversion of pyruvate to acetyl-CoA, high glucose activates ChREBP and induces the expression ACC and FAS, which facilitates the conversion of acetyl-CoA to lipids. During this process, ACC catalyzed
carboxylation of acetyl-CoA to malonyl-CoA suppresses fat oxidation through allosteric inhibition of CPT-1 [43]. In normal physiology, elevated glucose is always accompanied by high insulin, which suppresses the genes of fat oxidation and induces SREBP. SREBP, in turn, promotes the expression of lipogenic genes and enhances the conversion of acetyl-CoA to lipids [93].

**Summary**

The liver has been studied for its metabolic characteristics for more than 200 years. The liver’s ability to produce or utilize glucose is the foundation of glycemic control. During fasting, liver releases glucose from glycogen stores and uses gluconeogenesis to produce glucose from pyruvate, lactate, amino acids, and glycerol. During prolonged fasting, gluconeogenesis is upregulated to compensate depleted glycogen levels and prevent major changes in blood glucose concentration. Upon feeding, liver rapidly converts to an organ that utilizes more glucose than it produces. Postprandial glucose uptake by liver replenishes glycogen and provides substrate for lipid synthesis for long-term energy storage. A large proportion of gluconeogenesis is formed by “indirect” synthesis, a process driven by the simultaneous activity of glycolysis and gluconeogenesis. Lipids are synthesized from acetyl-CoA derived from the activation of glycolysis and PDH. These elegant modifications in metabolism are regulated by transcription, posttranslational modification, allosteric and substrate regulation of several key enzymes in the gluconeogenic and glycolytic pathway. Glucagon and other counterregulatory hormones activate transcription factors and co-activators that induce the expression and stability of gluconeogenic enzymes, while insulin opposes these effects and stimulates glucose uptake. Defects in any of these mechanisms can lead to loss of glycemic regulation.

**References**


CHAPTER 14

Insulin actions in vivo: glucose metabolism

Ele Ferrannini1 and Ralph A. DeFronzo2
1Department of Clinical and Experimental Medicine, University of Pisa, Pisa, Italy
2University of Texas, San Antonio, TX, USA; Health Science Center, San Antonio, TX, USA

Introduction

Glucose

Glucose is widespread in living organisms and, with protein and fat, completes the triad of the major metabolic fuels. To a much lesser extent than in plants, glucose also constitutes a building block for structural and enzymatic components of cells as well as the extracellular matrix. As a metabolic substrate, glucose is present in organisms essentially in its simple, monomeric form (α-D-glucopyranose) and as a branched polymer of α-glucose, namely glycogen. Disaccharides of glucose (lactose, maltose, and sucrose) are quantitatively less important. Glucose is present in plasma water at a concentration that—in a healthy adult who has fasted overnight—ranges between 3.6 and 5.5 mmol L⁻¹ (65–99 mg dL⁻¹). A family of proteins residing in the plasma membrane (and in microsomal membranes) can specifically and reversibly bind glucose molecules, and transfer them across cell membranes in both directions. Of such proteins, known as glucose transporters (or, more generally, Solute Carriers 2A or SLC2A), there are 14 different species that have so far been identified [1,2]. They differ from one another in both tissue distribution and physiologic regulation, particularly with respect to sensitivity to insulin stimulation. It is usual for more than one species of glucose transporter to be expressed in a tissue. One type of glucose transporter (GLUT4) represents the major insulin-stimulated glucose transporter in vivo, and this transporter is abundantly expressed in the classic insulin-sensitive tissues (adipocytes, brown fat, and skeletal, cardiac, and smooth muscle). Variable dominance of the other types of glucose transporters is found in tissues in which glucose metabolism does not respond to insulin acutely (erythrocyte, liver, kidney, brain, pancreatic β cells).

A non-insulin-regulatable transporter (GLUT1) effects facilitated glucose diffusion in red blood cells (RBCs). The abundance of this transporter in RBCs has the following physiologic consequences:

1. Glucose diffuses very rapidly across RBC membranes, with an estimated equilibration time of only 4 s (a total RBC mass of $25 \times 10^9$ cells with a mean diameter of 7 μm and a spherical shape occupies a surface area of approximately 4 m²) [3]. The rate of glycolytic utilization of transported glucose by the erythrocyte mass is estimated to be 25 μmol min⁻¹ or about $6 \mu \text{mol} \times \text{min}^{-1} \times \text{m}^{-2}$ of diffusion surface. Since this rate is approximately 17,000 times less than the rate of glucose transport in these cells, glucose concentration will be the same in plasma and erythrocyte water. Plasma proteins make up ~8% of plasma volume, while RBC proteins and ghost occupy about 38% of packed red cell volume (which, in turn, averages 40% of blood volume). Thus, 20% (i.e., $0.38 \times 0.4 + 0.08 \times 0.6 = 0.2$) of total blood volume is inaccessible to glucose. It follows that glucose concentration should be identical in plasma and RBC water under most circumstances, and that a blood water glucose concentration of 5.0 mmol L⁻¹ translates into a plasma glucose concentration of 4.6 mmol L⁻¹ and a whole-blood glucose concentration of 4.0 mmol L⁻¹, that is, a 15% systematic difference between plasma and whole-blood glucose level under typical conditions of hematocrit, proteinemia, and erythrocyte volume.

2. As both RBC and plasma convey glucose, the total amount of the sugar reaching any given organ is the product of arterial whole-blood glucose concentration times the total blood flow to that organ. Similarly, the total amount of glucose leaving a body region is the product of whole-blood glucose level in the venous effluent times the blood flow rate. Thus, under steady-state conditions of blood flow ($F$), arterial glycemia ($A$) and organ metabolism, the net balance of glucose movement across a body region is given by the product of blood flow and the arteriovenous ($A-V$) whole-blood glucose concentration.
Figure 14.1 Schematic representation of substrate (i.e., glucose) exchange across an organ (i.e., the liver) that both irreversibly removes the substrate (i.e., glucose) and adds it to the systemic circulation. A, arterial concentration; F, blood flow; V, venous concentration.

Net balance = Input - Output; U-R = F×(A-V)

Inflow (F × A)
Release (R)
Output = F × V + U

Uptake (U)

Net balance = Input - Output; U-R = F×(A-V)

Direct measurement of interstitial glucose concentrations has proven to be difficult, and has yielded conflicting results. Difficult as it may be to measure or calculate this “bathing” concentration of glucose, it is this concentration that dictates the activity of cellular glucose transport together with the state of specific activation of the glucose transporter. In addition to substrate mass action, cellular glucose uptake is influenced by changes in blood flow and hormonal stimulation by insulin. Thus, glucose metabolism is regulated by a distributed control between tissue glucose delivery (blood flow), transit through the interstitium, plasma membrane glucose transport and glucose phosphorylation (hexokinase) [6,7].

Blood flow varies considerably between insulin-sensitive tissues. Thus, basal blood flow to muscle and adipocytes in low (0.03–0.04 mL min⁻¹ × gram tissue) [8] and blood flow can play an important role in glucose delivery, cellular uptake, and subsequent metabolism. Baron [9] was amongst the first to show that physiologic hyperinsulinemia, while maintaining euglycemia, stimulated muscle blood flow resulting in enhanced muscle glucose uptake. The insulin-mediated increase in muscle blood flow results from two separate effects: (i) relaxation of resistance of vessels, and (ii) recruitment of previously unperfused muscle tissue secondary to relaxation of terminal arterioles [8]. This recruitment effect of insulin has been elegantly documented using microbubbles in combination with contrast-enhanced ultrasonography in rodents and humans [8]. The result is expansion of the capillary surface area for nutrient and insulin delivery. Impaired insulin-mediated vasodilation is associated with metabolic resistance to the stimulatory effect of insulin on muscle glucose utilization in type 2 diabetic and obese nondiabetic individuals [8,9]. Evidence in humans has demonstrated that the insulin concentration in lymph and interstitial fluid is considerably less than in plasma [10,11]. Transport rates of large molecules, such as insulin, into the interstitial space, and thus the cell membrane, is slow. Insulin binding to its receptor in the vascular endothelium and subsequent incorporation into caveolae are involved in the transcytosis of the hormone from the intravascular to the interstitial space [8]. Since the rate of insulin entry into skeletal muscle appears to be a critical step for insulin action, elucidation of the molecular and anatomical mechanisms involved in the transvascular process is likely to produce novel insights into the etiology of metabolic insulin resistance.

Another non-insulin-regulatable glucose transporter (GLUT2) is abundantly expressed in the plasma membrane of liver, kidney, intestinal cells, and pancreatic β cells [1,2]. In liver and kidney, net glucose release occurs in vivo. Thus, in the only organs in the body in which the presence of glucose-6-phosphatase (G6Pase)—the enzyme catalyzing the formation of free intracellular glucose from glucose-6-phosphate (G6P)—makes glucose available to the circulation, the transporter is of a type that only responds to the concentration gradient between the internal and external side of the plasma membrane. This ensures that, when insulin is around to stimulate inward glucose transport in tissues with sensitive transporters, the liver can release glucose into the blood stream as long as sufficient G6P is derived from glycogenolysis or gluconeogenesis (or both) and sufficient G6Pase activity is there to accumulate free glucose on the inside of the plasma membrane. Under these conditions of reversed gradient, the presence of insulin-responsive glucose transport...
activity on hepatocyte cell membranes would enhance glucose outflow, thereby opposing the plasma glucose-lowering action of insulin. In contrast, physiologic control of the direction and rate of glucose flux through the hepatocyte membrane is not on the transport step but on intracellular processes.

GLUT4 is the insulin-regulated transporter present in muscle and adipocytes [1,2]. In the fasting state, less than 5–10% of GLUT4 is present in the plasma membrane. The other 90% resides in GLUT4 storage vesicles (GSV) and endosomes within the cell. Following stimulation by insulin, the GSV translocate to the plasma membrane via exocytosis and mediate glucose transport into myocytes and adipocytes. Following binding of insulin to its receptor and activation of the insulin signaling pathway, a series of small (20–35 kDa) GTPases are activated and they interact with multiple motor proteins, membrane tethers, and fusion-regulating proteins to direct flow of the GSV to the plasma membrane [1,2]. GSV exocytosis can be divided into three separate processes: translocation to the cell periphery, targeting of the vesicles to the plasma membrane, and ultimately fusion with the cell membrane. In myocytes the phosphoinositol-3-kinase (PI3K) signaling pathway is an absolute requirement for GLUT4 exocytosis. In individuals with type 2 diabetes mellitus (T2DM) defects in both the PI3K signaling pathway and in the GLUT4 exocytotic pathway contribute to insulin resistance in muscle and adipocytes [12]. Following exposure to insulin, GLUT4-containing vesicles can be demonstrated in the cell periphery and plasma membrane within 5 minutes. Disruption of the microtubular/actin system within myocytes/adipocytes completely inhibits insulin-stimulated GLUT4 translocation. Fusion of GSV with the plasma membrane requires interaction with SNARE (soluble N-ethylmolemide-sensitive factor attachment receptor regulatory) proteins including VAMP2, STX4, SNAP23, and others. In adipocytes a PI3K-independent signaling pathway mediated via the adaptor protein APS (adaptor protein with pleckstrin homology and Src homology domains) also plays an important role in GLUT4 exocytosis; in myocytes a role for the insulin-stimulated APS pathways has yet to be defined [2].

In summary, the differential distribution and acute insulin sensitivity of the various classes of glucose transporters provide the backbone for the functional characteristics of glucose diffusion and exchange in the body. On the whole, free glucose is present in blood water, interstitial fluid, and the intracellular water compartment of insulin-independent tissues (liver, brain, kidney, intestine, placenta) in total amounts which, in the overnight-fasted healthy adult, average 80 mmol (14 g or 1.2 mmol·kg$^{-1}$ of body weight), of which one fifth is in the blood volume. Free glucose is found at concentrations that (a) are uniform in the intravascular water compartment; (b) decline across the interstitial space toward the cell; (c) fall precipitously within cells that consume glucose avidly (e.g., brain) or in which glucose transport is relatively slow in the basal state (e.g., muscle, adipose tissue); (d) are increased in the cytoplasm of cells that produce free glucose (mostly liver); and (e) gradually and continuously decrease in the vascular bed as arterial blood turns into capillary blood and then runs back toward the right heart as venous blood. The regional characteristics of tissue composition, blood flow rate, capillary density (i.e., the average distance between the capillary axis and the cell surface), and cellular glucose uptake concur to determine the A-V glucose gradient in any region of the body. Glycogen is present in most cells in cytoplasmic granules that encase the enzymes that regulate its metabolism. In normal humans, the largest part of glycogen stores is in liver and skeletal muscle. In the former, 3–4 g of glycogen are packed in each 100 g of parenchyma; in striated muscle, the concentration is much lower (0.7–1.0% weight by weight). As a consequence, a normal human liver (1.5 kg) contains some 60 g of glycogen, whereas muscle (28 kg) depots keep 250 g. Thus, approximately 25 times more glycosyl units are stored in intracellular depots as glycogen than are dissolved in body water as free glucose. Glycogen metabolism is controlled by irreversible cascades of enzymatic reactions ultimately acting upon the proximal enzymes that catalyze glycogen synthesis (glycogen synthase) and degradation (glycogen phosphorylase) (see subsequent discussion).

**Key points**
- Glucose concentration is on average 15% higher in the plasma than the whole-blood glucose under ordinary circumstances.
- Arterial glucose levels are higher than capillary glucose concentrations, which in turn are higher than venous levels. The extent of these differences depends on the circulatory region and on its rate of glucose utilization.
- Both plasma and erythrocytes carry glucose to and from consuming tissues.
- In brain, liver, kidney, intestine, and placenta, glucose uptake is insulin independent; in adipose tissue, skeletal and heart muscle, glucose uptake depends on insulin.
- Tissue-specific glucose transporters (GLUTs) mediate glucose uptake in insulin-sensitive and insulin-insensitive tissues.
- The whole-body reserves of carbohydrate are quite small.
- Insulin-mediated vasodilation, involving recruitment of unperfused capillary beds, plays a role in regulating muscle glucose uptake.

**Glucose metabolism**

**Introduction**

Any given concentration of glucose in plasma (and in the space in equilibrium with plasma) is the result of simultaneous release of glucose into the circulation and uptake of glucose from the blood stream into cells. Whenever plasma glucose concentration is stable, the concurrent rates of its release and overall uptake must be equal. Glucose turnover (TR) is this rate of constant flux through its system. Accepted terminology...
refers to stationary conditions of glucose concentrations and flux rate as the “steady state,” with the understanding that any physiologic steady state is only an approximation of the true, ideal steady state. Furthermore, the “glucose system” refers to the whole space or volume into which glucose is present as free glucose, regardless of how many compartments this system consists of, where they are physically located in the body, and how they are interconnected. Finally, the reference pool for glucose kinetics customarily is the plasma (or whole blood), for in most clinical or experimental circumstances the plasma is the only site accessible for sampling. When the plasma glucose concentration changes over time, one rate of glucose flux (entry or removal) is being exceeded by the other. Under these non-steady-state conditions, the glucose rates of entry and removal are conventionally termed rate of appearance (Ra) and disappearance (Rd), respectively. The glucose system is strongly homeostatic with respect to glucose levels, in that the normal variations in human plasma glucose concentration throughout a day of life are confined within a surprisingly narrow range. It should be recalled that glucosuria is called upon whenever glycaemia exceeds the renal threshold, $\sim 180$ mg dL$^{-1}$, as if a safety measure had been set to cope with emergency when metabolic control fails. These considerations alone indicate that the body does not tolerate either hypoglycemia or hyperglycemia. For the former, the obligate dependence of brain function on the use of glucose as fuel classically has been offered as a rational explanation. For hyperglycemia, the evidence—and hence the concept—that high glucose levels, if not immediately life-threatening, are nonetheless intolerable to bodily functions, is more recent but no less compelling, and currently goes under the name of “glucose toxicity” (see Chapter 27).

**Methods**
Under conditions of an overnight fast, the liver accounts for $\sim 80\%$ of glucose production with the remaining $\sim 20\%$ coming from the kidney [12,13]. Placing catheters across the splanchnic bed (one in a hepatic vein and another in any artery) and measuring splanchnic blood flow (e.g. by infusing a dye, such as indocyanine green, that is only cleared by the hepatocyte) allows the measurement of glucose turnover as the product of $A-V$ difference and blood flow (Figure 14.1). However, the liver and extrahepatic splanchnic tissues (gut, pancreas, spleen, etc.) also take up glucose; the application of the Fick principle allows one to estimate the net balance between glucose uptake and release in the splanchnic area, not the total rate of glucose turnover (Figure 14.1). One must resort to glucose tracers, such as radioactive (e.g. $^3$H-glucose or $^{14}$C-glucose) or stable (e.g. $^2$H-glucose or $^{13}$C-glucose) isotopes of glucose. The kidney, like the liver, also simultaneously takes up and releases glucose. By combining tracers and renal vein catheterization with measurement of renal plasma flow (using para-aminohippurate), one can quantitate the bidirectional flux of glucose across the kidney, similar to that described across the splanchnic bed. Details of the tracer technique as applied to glucose turnover measurement can be found in several reviews and treatises [14,15].

A tracer can be administrated as a pulse injection or constant intravenous infusion or combination thereof. Under steady-state conditions applying to both tracee (i.e., glucose) and tracer, glucose turnover rate is simply given by the ratio of the tracer infusion rate (IR) to the equilibrium plasma glucose specific activity (SA = tracer/tracee concentration). Equilibrium is the time (usually 2–3 h after starting the tracer infusion) when unchanging plasma tracer concentrations indicate that glucose specific activity has become uniform throughout its distribution space. This formula, IR/SA, is not based on any assumptions other than the attainment of equilibrium. When non-steady-state conditions prevail, this approach cannot be used because either the tracer or the tracee concentration (or both) change over the period of measurement. Unfortunately, in the patient with diabetes the glucose system is unsteady for most of the time, precisely because homeostasis has been lost. However, practicable ways around the problem do exist [16] and the reader is referred to several reviews of the topic [14,15].

As discussed earlier, glucose tracers are useful in the study of regional metabolism in two respects:

1. For organs in which glucose uptake and release occur simultaneously (typically, the liver and kidney) the $A-V$ difference for the tracer offers a measure of absolute uptake (for the tracer is not produced by the organ), while absolute release can be calculated as the difference between net balance (as measured by the Fick principle) and uptake. If $A^*$ and $V^*$ and $A$ and $V$ are the arterial and venous tracer and tracee concentrations, respectively, and $F$ is the blood flow, the absolute glucose uptake rate is $(GU) = (A^*-V^*)/A \times A \times F$ and the net balance is $(NB) = (A-V) \times F$. The rate of glucose production is $(GP) = NB - GU = FA \times A \times (V^*/A^*-V/A)$.

2. As tracer glucose is metabolized, its label will appear in one or more degradation products $(P^*)$. For any degradation product, the ratio of its labeled moiety to the specific activity of glucose (i.e., $P^* 	imes (A^*/A)^{-1}$) is the amount of that product generated from glucose (in units of concentration). For example, measuring $^{14}$C-lactate and $^{14}$C-carbon dioxide during the infusion of $^{14}$C-glucose makes it possible to estimate the amounts of glucose that were glycolyzed and completely oxidized, respectively. If, then, the kinetics of precursor and product are separately determined, all of these precursor–product relationships can be converted into fluxes of interconversion, regionally as well as at the whole-body level.

An important concept in kinetics and physiology is that of glucose clearance. When referred to the plasma volume, glucose clearance is the volume of plasma that is completely cleared of glucose per unit time. Glucose clearance (CR) is related to glucose turnover rate as follows: $TR = CR \times G$ (where $G$ is the glucose concentration in the same pool, plasma or blood, arterial or venous, in which the tracer concentration was determined). At the level of an organ that only consumes glucose—
Glucose turnover can occur in other tissues (e.g., skeletal muscle [17]), but the absence of significant G6Pase activity in tissues other than the liver and kidney prevents labeled breakdown products from re-entering the circulation as glucose, where they would alter the estimation of glucose turnover. As for tritiated or deuterated isotopes, labels in positions 3, 4, 5, and 6 are lost at the triosephosphate step or further downstream in anaerobic glycolysis, while a label in position 2 is largely lost at the phosphoglucoisomerase step (a near-equilibrium reaction) soon after phosphorylation [14,15,18] (Figure 14.2). In neither case does the detached label (essentially in equilibrium with the hydrogen of the body water pool) recycle back into a new glucose molecule to any detectable extent.

In the basal state, glucose output in healthy adults averages \( \sim 840 \mu\text{mol min}^{-1} \) (or \( \sim 12 \mu\text{mol min}^{-1} \text{ per kg of body weight} \)) [19]. The dispersion around this mean estimate is significant (20–30%), with an unknown contribution of genetic and environmental factors. In nondiabetic subjects total body endogenous glucose output variability is wide and is largely explained by the amount of lean mass [20] and this, in turn, explains differences in total endogenous glucose output due to sex, obesity, and age. Under standard nutritional conditions, the fasting liver depletes its glycogen stores at a rate of about 5% per hour, such that glycogen depots are empty after 24 h. Since fasting can be prolonged well beyond 24 h, obviously gluconeogenesis must progressively replace glycogenolysis as fast continues [21,22].

In animal species in which the basal rate of glucose turnover is higher than in humans (e.g., dogs, 20 \( \mu\text{mol min}^{-1} \text{ kg}^{-1} \); rats, 40 \( \mu\text{mol min}^{-1} \text{ kg}^{-1} \)), the limited capacity of the liver to store glycogen confers an increasing role to gluconeogenesis for the maintenance of basal glycemia. This limitation on glycogen accumulation has an anatomical basis: overcrowding of cytoplasm with glycogen granules impairs cellular functions, leading to infiltration of nuclei and, eventually, to cell death, as observed in several glycogen storage diseases. In normal healthy adults in the overnight fasted state, gluconeogenesis and glycogenolysis approximately equally contribute to hepatic glucose release [21–23]. In obese nondiabetic and in type 2 diabetic subjects the contribution of gluconeogenesis to hepatic glucose production is increased compared to lean normal glucose tolerant subjects [22,24,25]. In subjects with variable degrees of overweight and hyperglycemia, it has been established that the percent contribution of gluconeogenesis to fasting glucose release rises with increasing body mass index (by \( \sim 1\% \) per body mass index unit) and increasing fasting hyperglycemia (by \( \sim 3\% \) per mmol L\(^{-1} \)) [22]. In healthy subjects, physiologic hyperinsulinemia suppresses percent gluconeogenesis by \( \sim 20\% \) while completely blocking glycogenolysis [24,26]. As long as hyperinsulinemia restrains glycogenolysis—as is the case of obese nondiabetic subjects—basal endogenous glucose release will be normal in absolute terms. As glycogenolysis also becomes resistant to the inhibitory effect of insulin—as in the case of type 2 diabetic patients [26]—fasting glucose output
Chapter 14

−μ−

Splanchnic glucose uptake

μ kg

and hepatic (i.e., kg)

inhibitory (on the left) and stimulatory (on the right) factors

in healthy subjects under four experimental conditions: (i) overnight fast; (ii) hyperglycemia (+125 mg·dL−1) with somatostatin blockade of endogenous insulin release (Low insulin); hyperglycemia (+125 mg·dL−1) with somatostatin plus insulin replacement to maintain the fasting insulin concentration constant (Basal insulin); hyperglycemia (+125 mg·dL−1) with endogenous insulin (55 uU·mL−1) release. Note that hyperglycemia per se inhibits hepatic glucose production and that hyperglycemia acts synergistically with insulin to inhibit liver glucose output. In contrast, hyperglycemia stimulates glucose uptake to approximately the same extent in the presence of low, basal, or high insulin, that is, mostly by mass action. Source: Drawn from the data presented in DeFronzo et al. 1983 [32].

As for metabolic signals, hyperglycemia per se can effectively inhibit liver glucose output in normal humans [29]. As shown in Figure 14.4, during constant hyperglycemia (maintained by the hyperglycemic clamp technique) hepatic glucose production is significantly reduced in comparison with euglycemia even when the endogenous insulin response to hyperglycemia is blocked by somatostatin. It is evident from Figure 14.4 that normally the two inhibitory signals, hyperglycemia and endogenous hyperinsulinemia, concur to shut off glucose production [32–34]. Hypoglycemia by itself may trigger an increase in hepatic glucose release even when one or more counterregulatory influences are paralyzed [35]. Studies in humans [33,36] and in dog [37] have established a role for a portosystemic glucose concentration gradient in controlling hepatic glucose handling. Thus, intraportal glucose infusion specifically alters the partitioning of a glucose load among tissues, by stimulating net hepatic glucose uptake and reducing peripheral glucose uptake. Of other metabolic signals, it is difficult to demonstrate in humans an increase in hepatic glucose production by infusing large quantities of glycerol, lactate, or a mixture of amino acids, as long as there is physiologic hyperinsulinemia to balance out such a gluconeogenic push.

Note, that an increased provision of precursors may well lead to increased intrahepatic formation of glucose-6-phosphate (G6P), but the eventual fate of this intermediate may be glycogen rather than free glucose if control of the rate-limiting step for free glucose production, namely G6Pase activity, is not simultaneously loosened. Free fatty acids (FFA) stand out as one substrate that may play an important role in setting the level of hepatic glucose production. Since only odd-chain FFA...
(i.e., propionate) can donate their carbon to oxaloacetate in the tricarboxylic acid cycle, FFA do not represent an important precursor for gluconeogenesis. Most physiologic FFA are even-chain, and they exchange their carbon moieties with tricarboxylic acid cycle intermediates but do not contribute to de novo glucose synthesis. Nevertheless, enriching the perfusion medium of isolated rat liver with oleate or palmitate [38]. Moreover, FFA and/or products of their oxidation (e.g. citrate and acetyl-CoA) activate key gluconeogenic enzymes such as pyruvate carboxylase, phosphoenol-pyruvate carboxykinase, and G6Pase [38,39]. In addition, raised FFA concentrations in vivo are accompanied by raised glycerol levels, resulting from hydrolysis of triglycerides. Therefore, accelerated lipolysis normally supplies both the stimulus (FFA) and the substrate (glycerol) for gluconeogenesis. Finally, the liver takes up FFA avidly (with an extraction ratio of ~30%), and oxidizes them efficiently (as indicated by the low respiratory quotient of the organ) [40,41]. Thus, there are all the requisites to consider FFA oxidation in the liver as the energy-providing process that is coupled to energy-requiring gluconeogenesis. In isolated hepatocytes, FFA in micromolar amounts inhibit glycogen synthase [42], which suggests that an additional interaction of this substrate with hepatic glucose metabolism may be at the level of glycogen metabolism. In healthy volunteers, short-term infusion of triglycerides with heparin (to activate lipoprotein lipase and elevate the plasma FFA concentration) results in an increase in hepatic glucose output under conditions (hyperglycemia and somatostatin block of endogenous insulin response) mimicking the key features of diabetes [43]. When endogenous insulin is allowed to rise, or when exogenous insulin is administered, the stimulatory effect of triglyceride infusion on hepatic glucose release is easily overcome. Consistent with this, FFA infusion in healthy subjects augments hepatic gluconeogenesis but basal hepatic glucose production does not rise because of a reciprocal decrease in hepatic glycogenolysis [40,44].

In summary, long-chain FFA may regulate glucose production both by acting upon key enzymes of gluconeogenesis (through products of FFA oxidation) and by virtue of the substrate push of glycerol. In glucose-tolerant normal subjects, this regulatory mechanism is primarily operative when insulin secretion is not stimulated, that is, in the basal state. Insulin and the classic counterregulatory hormones—glucagon, cortisol, growth hormone (GH), catecholamines, and triiodothyronine (T₃)—form one of the best-described agonist–antagonist regulatory systems (Figure 14.3). Insulin is a potent, specific, and rapid-acting inhibitor of hepatic glucose production [12,45]. It restricts both glycogenolysis and gluconeogenesis, although with different dose–response characteristics, gluconeogenesis being less sensitive [24,46]. Moreover, by restraining lipolysis and proteolysis, insulin also reduces the delivery of potential glucose precursors (glycerol, amino acids) to the liver. In its capacity as the inhibitory signal for glucose release, insulin is greatly favored by the anatomical connection between the pancreas and the liver, and secreted insulin reaches the liver at a concentration that in fasting humans is three- to fourfold higher than the peripheral (arterial) concentration [47]. Such portosystemic gradient is maintained by a high rate of insulin degradation by hepatic tissues (with a fractional extraction of about 50%). Thus, a small secretory stimulus to the β cell primarily serves to increase portal insulin levels, thereby selectively acting upon glucose production rather than also enhancing peripheral glucose utilization. In addition to short-circuiting the systemic circulation, pancreatic insulin release is potentiated by several gastrointestinal hormones (e.g. glucose-dependent insulinoctropic polypeptide [gastric inhibitory polypeptide] and glucagon-like peptide 1). Therefore, anatomical and physiologic connections in the gut–liver–pancreas circle ensure that the primary station for the handling of foodstuff, the liver, is under close control by a nearby, well-informed unit, the β cell.

The anti-insulin hormones all counter insulin action on the liver by facilitating both glycolgenolysis and gluconeogenesis. They do so, however, with different dose–response kinetics and time courses. Thus, glucagon and catecholamines act rapidly, while cortisol, growth hormone, and thyroid hormones (in that order) are involved in the long-term control of glucose release [48]. Glucagon plays a major part in the tonic support of hepatic glucose release: in humans suppression of glucagon release with preservation of basal insulin secretion causes a fall of glucose production of over one third [48]. The precise quantitative contribution of the other counterregulatory hormones system to the maintenance of basal glucose output has not been assessed. What has become established, however, is the strongly synergistic pattern of interaction between the anti-insulin hormones, such that their cumulative effect is larger than the sum of the individual effects [49,50]. This interaction also is expressed at the level of the components of glucose production, glycogenolysis and gluconeogenesis, as well as at the level of peripheral tissues (as discussed later in this chapter). An added feature of this homeostatic complex is the dual negative feedback between agonists and antagonists (Figure 14.3). Alpha cells and β cells “talk” to each other in the islets of Langerhans via paracrine influences (some of which may possibly be mediated by somatostatin). For example, insulin infusion in vivo reduces circulating glucagon levels by 20–30% even when changes in glycemia are prevented [51].

Another example is the direct modulation of β-cell secretion by catecholamines, with α-adrenergic stimulation (mostly, α-2) inhibiting and β-adrenergic stimulation increasing insulin secretion. It is noteworthy that in this control system a host of hormones is required to balance the action of only one agonist, insulin. This fact arises from the inhibitory nature of insulin’s effect on the production of a fuel upon which brain cell viability depends in an obligatory manner.
Key points

- After a subject has fasted overnight, all of the glucose entering the circulation (at a mean rate of \(12 \mu\text{mol min}^{-1} \text{kg}^{-1}\), or about 220 g day\(^{-1}\)) derives from liver glycogenolysis and liver/renal gluconeogenesis, with approximately equal contributions from each metabolic pathway.
- Approximately 80% of endogenous glucose production is derived from the liver and 20% from the kidney.
- As fasting continues, glucose production becomes progressively more dependent on gluconeogenesis.
- Any given rate of glucose output is the net result of the inhibitory actions of insulin, hyperglycemia, parasympathetic nervous activity and substrate shortage on the one hand, and the stimulatory actions of counterregulatory hormones, hypoglycemia, sympathetic nervous activity, and gluconeogenic substrate load on the other.
- Among substrates, FFA have an added regulatory value for glucose output.
- Inappropriately elevated plasma insulin levels can be life-threatening in the fasting state by shutting off endogenous glucose supply to the brain. Insulin counterregulation, therefore, is entrusted to several hormones, providing both short-term and long-term control.

Glucose disposal

In the basal state, steady/near steady-state conditions prevail and whole-body glucose disposal equals endogenous (hepatic) plus renal glucose production. Data on the individual contribution of organs and tissues to total glucose uptake have been obtained in regional catheterization studies. In the case of the splanchnic area, in which glucose uptake and production both occur simultaneously, such data have been derived from the combined use of glucose tracers and indwelling catheters, as diagrammatically shown in Figure 14.1. By collating the available information, the organ-circulation model of Figure 14.5 can be drawn, in which steady-state inter-organ exchanges of blood and glucose, and regional glucose gradients, are calibrated at a rate of endogenous glucose release of 840 \(\mu\text{mol min}^{-1}\) (or 12 \(\mu\text{mol min}^{-1} \text{kg}^{-1}\) or 1.2 mol day\(^{-1}\)). In this model, it is seen that roughly 70% of basal glucose disposal takes place in insulin-independent tissues (brain, liver, kidney, intestine, erythrocytes). It can also be appreciated that the fractional extraction of glucose is quite low everywhere in the body (ranging from 1.7 to 2.8%) except in the brain (9%). If it is assumed that skeletal muscle (40% of body weight) receives 16% of cardiac output and is responsible for one quarter of overall glucose disposal (~245 \(\mu\text{mol min}^{-1}\); Figure 14.5), then muscle glucose clearance averages 1.3 mL min\(^{-1} \text{kg}^{-1}\) of tissue. This value can be compared with those of other organs and tissues (Table 14.1) similarly obtained by dividing the organ glucose clearance by the estimated organ weight. In the rank of efficiency of glucose utilization in the basal state, resting muscle is last, being 10 times less active than the liver, and 50 times less avid than the brain. In tissues in which specific glucose clearance is already high (brain, liver, kidneys), the effect on glucose uptake of acutely raising plasma insulin levels above fasting values is small or absent, while in muscle glucose clearance can increase by a factor of 10 over the physiologic range of insulin concentrations. The intermediate position of heart muscle in the list is likely accounted for by its working state. As reviewed earlier, these

Figure 14.5 Schematic representation of organ glucose metabolism and blood flow in the basal or postabsorptive state. Average data compiled for healthy adults from the literature are indicated. “Periphery” encompasses all tissues other than liver, gut, kidneys, brain, and heart; “gut” includes organs (i.e., spleen, pancreas, etc.) draining their blood supply into the portal circulation. CA, carotid arteries; HA, hepatic artery; HV, hepatic vein; MA, mesenteric arteries; PV, portal vein; RV, renal veins; VA, vertebral artery. Organ blood flow is shown in mL min\(^{-1}\) and glucose fluxes in mg min\(^{-1}\) (\(\mu\text{mol min}^{-1}\)).
characteristics can now be seen as the physiologic equivalent of the type and abundance of specific glucose transporters with which the various tissues are endowed. They also narrow the concept of an insulin-independent tissue: thus, raising the plasma insulin concentration does not accelerate glucose clearance. However, lowering insulin may reduce the efficiency of glucose removal, and even non-insulin-regulatable glucose transporters are subject to chronic regulation. In humans this has been verified for at least two insulin-independent tissues: the liver and the brain. For the former, selective hypoinsulinemia (induced with somatostatin and glucagon replacement) lowers splanchnic glucose uptake below the basal value in healthy volunteers [32] (Figure 14.4). For the latter, indirect calculations indicate that brain glucose clearance is reduced in type 2 diabetic patients with moderate fasting hyperglycemia [52]. The intracellular disposition of transported glucose can be studied by using glucose tracers, and then tracking down the appearance of the label in specific metabolic products, such as lactate (i.e., anaerobic glycolysis) or carbon dioxide (i.e., complete oxidation) (see earlier discussion). It should be noted that these techniques, even when correctly applied, provide estimates of the metabolic fate of plasma glucose, which is the labeled pool. If, for example, there should be direct oxidation of glycogen in muscle, the plasma glucose specific activity would miss it (because of the lack of G6Pase in muscle tissue). However, total carbon dioxide production, measured with indirect calorimetry, would include the oxidized glycogen together with all the other oxidized substrates. In general, measuring the exchange of oxygen and carbon dioxide with indirect calorimetry makes it possible to obtain estimates of net rates of substrate oxidation and also provides a close estimate of the rate of energy expenditure [53]. In the basal state and under ordinary nutritional circumstances, oxygen consumption averages ∼250 mL min⁻¹, while carbon dioxide production is ∼200 mL min⁻¹, that is, a whole body respiratory quotient of 0.8 (RQ = carbon dioxide production/oxygen consumption). Simple calculations [54] thus estimate whole-body net carbohydrate oxidation at about 60% of total glucose uptake. As the brain uses 46% of glucose turnover (Table 14.1) and readily oxidizes the transported sugar, it follows that three quarters (i.e., 46/60 = 77%) of basal glucose oxidation occurs in the brain. Little is left for other tissues, which preferentially derive their metabolic energy from the oxidation of fatty substrates and return most of the glucose to the liver after conversion into lactate (Cori cycle). Skeletal muscle, for example, has a respiratory quotient of 0.75 and relies on fat oxidation for the production of 80% of the energy it needs in the resting state [55]. Thus, the basal state is characterized by parsimonious usage of glucose as fuel, which is selectively channeled to organs that cannot rely on alternative energy sources. Altogether over one half of total energy production (5 kJ min⁻¹) is generated via oxidation of fat, of which there are plentiful (∼500 MJ) and almost unlimited stores (e.g. obesity). The role of insulin in maintaining this metabolic setup is permissive rather than determinant. Through a loose brake on lipolysis, insulin lets FFA override glucose in the competition between the two chief substrates, keeps glucose transport and metabolism in its own target tissues at a minimum, and restrains protein breakdown (which contributes only about 15% to energy metabolism). The part that counterregulation plays in basal glucose uptake is less well defined, but probably is centered upon maintenance of lipolysis, since all the anti-insulin hormones are more or less potent lipolytic stimuli.

### Glucose cycles

After transport through the plasma membrane, glucose does not necessarily follow a straight path to its eventual fate—be it glycogen, pyruvate, or pentoses—but may go around in what are known as futile cycles. A metabolic futile cycle is one in which a precursor is converted into a product by a forward reaction, which is then reversed to resynthesize the precursor. In this way, no net product accumulates, but energy (ATP) is used. One example of a futile cycle, G6P to fructose-6-phosphate and back through the phosphoglucoisomerase reaction, has been discussed previously and can be measured as the difference between the turnover rate as estimated by ²⁵H and ³H isotopes of glucose. Another example is glucose to G6P via glucokinase and back via G6Pase. In general terms, whenever bidirectional flux through a metabolic pathway is simultaneously operative, there exists a cycle, regardless of the number of intermediate reactions and regardless of whether one or more tissues are involved. In this sense, lipolysis in adipose tissue, followed by partial

### Table 14.1 Regional glucose disposal in the basal state

<table>
<thead>
<tr>
<th>Organ</th>
<th>Weight (kg)</th>
<th>Blood flow (l min⁻¹)</th>
<th>Uptake (µmol min⁻¹)</th>
<th>Extraction (%)</th>
<th>Clearance⁺ (mL min⁻¹ kg⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brain</td>
<td>1.2</td>
<td>0.85</td>
<td>385</td>
<td>9.1</td>
<td>64</td>
</tr>
<tr>
<td>Liver</td>
<td>1.5</td>
<td>1.50</td>
<td>110</td>
<td>2.3</td>
<td>15</td>
</tr>
<tr>
<td>Kidneys</td>
<td>0.28</td>
<td>1.10</td>
<td>20</td>
<td>1.9</td>
<td>15</td>
</tr>
<tr>
<td>Heart</td>
<td>0.3</td>
<td>0.25</td>
<td>20</td>
<td>1.7</td>
<td>13</td>
</tr>
<tr>
<td>Gut</td>
<td>5.0</td>
<td>1.20</td>
<td>60</td>
<td>1.0</td>
<td>2.4</td>
</tr>
<tr>
<td>Muscle</td>
<td>28.0</td>
<td>1.05</td>
<td>245</td>
<td>3.5</td>
<td>1.3</td>
</tr>
</tbody>
</table>

⁺Organ clearance rate divided by organ weight.
reesterification of FFA in the liver, is a complete cycle. Another one is protein breakdown in skeletal muscle with reincorporation of amino acids into proteins in the liver. The negative connotation of futility has traditionally been reserved for those cycles that go on in the same cell. They are, however, anything but futile. As discussed by Newsholme [56], a metabolic cycle with an internal loop is the best kinetic stratagem to keep the enzymes of a dormant pathway at a minimum of activity, and to ensure a sensitivity gain for ready amplification of incoming signals. The ATP cost of these cycles is itself a means of increasing the efficiency of energy dissipation. The fact that the activity of these cycles is under hormonal control (e.g. catecholamines and thyroid hormones enhance the cycling rate) makes room for modulation; in this way, these cycles become components of facultative thermogenesis.

Key points
- Of fasting glucose uptake, 70% occurs in insulin-independent tissues, and two thirds of the glucose is completely oxidized. Much of this distribution is due to the obligatory use of glucose as fuel by the brain.
- In insulin-independent tissues, raising insulin does not stimulate glucose metabolism, but insulin lack may decrease the efficiency of glucose utilization (glucose clearance).
- Fat is the preferred substrate of insulin-dependent tissues in the basal state.
- Over 50% of basal energy production relies on fat oxidation.

The fed state

Introduction
The fed state is the absorptive period between meals. Carbohydrates are normally mixed with lipids and protein in the diet and make up 40–60% of the caloric content. Absorption of dietary carbohydrates is influenced by their chemical form (refined sugars or complex carbohydrates) and by other components of food. Furthermore, disposition of dietary carbohydrate is indirectly affected by fats and protein to the extent that these latter (i) compete with glucose as substrates, and (ii) interfere with glucoregulatory hormones by altering insulin secretion. To circumvent the difficulties of study of glucose ingestion, the regulation of glucose homeostasis during the fed state has classically been investigated with the use of intravenous glucose, which can be administered in formats more suitable for formal analysis.

Methods
Glucose can be injected intravenously as a single bolus (0.33 or 0.5 g [1.8 or 2.8 mmol] per kg of body weight) and the decline in plasma glucose concentration after the initial peak followed for 60–90 min (intravenous glucose tolerance test or IVGTT). Between 10–60 min, glycemia decreases approximately as a single exponential function of time, and a decay constant ($k$ value) can be calculated to estimate tolerance to intravenous glucose. The IVGTT has several drawbacks. First, the time course of glucose fall in reality is a multi-exponential function of time, so that the decay constant takes on different values according to which segment of the curve is used for analysis. Second, over the same time interval, different curves can have similar $k$ values; for this reason, the area under the glycemic excursion is sometimes used instead of the $k$ value. However, the area under the curve is also influenced by the volume into which the injected glucose is distributed. Third, and most important, the shape of the glucose curve is heavily affected by the endogenous insulin response to the acute hyperglycemia caused by the intravenous bolus. Such response is highly irregular, and after an initial peak proceeds in two or three smaller spikes tightly synchronized with similar glycemic spikes [57].

In general, in the presence of an intact feedback loop between glucose and insulin, glucose tolerance is the integrated outcome of multiple changes in the glucose as well as in the insulin system: distribution of the exogenous glucose, stimulation of peripheral glucose uptake by insulin and hyperglycemia, suppression of hepatic glucose output, and secretion, distribution and degradation of insulin. It is therefore not surprising that the information provided by an IVGTT is somewhat ambiguous, and that the test itself is neither very sensitive nor easily reproducible. A modern version of the IVGTT is that which interprets the changes in glucose and peripheral insulin concentrations (measured at frequent intervals following the bolus) on the basis of a “minimal” mathematical model of the glucose and insulin systems and their interactions [58]. The model generates a parameter reflecting the ability of hyperglycemia to stimulate insulin secretion, another parameter estimating the ability of insulin to stimulate glucose metabolism, and an index of the ability of glucose to promote its own disposal. Although attractive for its simplicity, the minimal model approach generally falls short of its ambitious goal, that is, to describe all aspects of glucose tolerance with a minimum of data. In particular, the minimal model does not work when the endogenous insulin response is scarce, for example in diabetic patients. The model is not minimal in that the data analysis requires a computer program that, like all package deals, deprives users of critical evaluation. Several updates of the technique have been devised. Labeled (tritiated or deuterated) glucose can be co-injected with cold glucose; appropriate model analysis of the tracer data makes it possible to dissect out the effects of insulin on the liver and on peripheral tissues [59]. Because of its complexity of use and interpretation, the latter remains a purely investigative tool.

A secondary injection of exogenous insulin or tolbutamide has been used to circumvent the failure of the minimal model in case of insufficient endogenous insulin response [60]. A simple way to estimate whole-body sensitivity to insulin is to paralyze endogenous insulin release with a constant infusion of somatostatin (at a rate of 0.3–0.5 mg h$^{-1}$) while simultaneously infusing glucose (at a rate of 1.35 mmol min$^{-1}$ m$^{-2}$) and regular
Figure 14.6 (top) Schematic representation of the euglycemic insulin clamp technique. See text for a detailed discussion. (bottom) Schematic representation of the hyperglycemic clamp technique. See text for a detailed discussion.

Insulin (at a rate of 50 μU min⁻¹ m⁻²). With this technique (also known as the pancreatic suppression test) [61], steady hyperinsulinemia (~80 mU L⁻¹) is associated with a level of hyperglycemia that is inversely proportional to the ability of whole-body tissues to increase their glucose utilization in response to insulin. This test suffers from the fact that somatostatin inhibits the release of other glycoactive hormones (e.g., glucagon). With this limitation, however, the test is simple and reliable enough for clinical use. The glucose clamp technique has become the reference method to study glucose metabolism [62]. Figure 14.6 (top) exemplifies the euglycemic hyperinsulinemic version of the clamp technique. An exogenous infusion of regular insulin is started at time zero in a format comprising a prime followed by a constant infusion (usually at a rate of 1 mU min⁻¹ kg⁻¹); such infusion quickly establishes a hyperinsulinemic plateau of about 60–70 mU L⁻¹. A few minutes after starting the insulin infusion, an infusion of glucose is begun at a rate which is adjusted every 5–10 min on the basis of on-line plasma glucose measurements obtained with the same frequency. Over the second hour of a 2-h experiment, euglycemia in the absence of constant hyperinsulinemia is maintained by an approximately constant glucose infusion, which in a healthy adult ranges between 20 and 50 μmol min⁻¹ kg⁻¹ (a mean value is shown in the figure). Such a rate equals the overall rate of glucose uptake (also called M for metabolism) in a subject in whom endogenous glucose production is nil. Relative insulin insensitivity or insulin resistance is indicated by a low M value at comparable levels of glycemia and insulinemia. The technique has the following advantages: (a) any preset combination of plasma glucose and insulin levels can be easily achieved; (b) the time course of insulin action can be determined with a time resolution of about 10 min; (c) hypoglycemia with its attendant counterregulatory hormonal response can be avoided; (d) other techniques, such as tracer glucose infusion, muscle biopsy, and indirect calorimetry, can be readily combined with a clamp protocol; (e) the interference of other hormones or substances with insulin action can be quantitated by co-infusing them during a clamp study; (f) high reproducibility.

Although computerized algorithms are available to run a clamp, manual operation with a minimum of experience does just as well. The major drawback of the euglycemic insulin clamp is the need to draw frequent blood samples from an arterialized vein (e.g. a heated wrist or hand vein). The hyperglycemic version of the glucose clamp (schematized in Figure 14.6, bottom) consists of acutely raising plasma glucose to any desired level, and then clamping it at that level with a variable infusion of exogenous glucose [62] (as in the euglycemic clamp version). The hyperglycemic step evokes an endogenous insulin response that typically is biphasic: an early output (of preformed hormone) that lasts 10–15 min, followed by a gradual, continuous rise in insulin levels (I), reflecting glucose-induced triggering and potentiation of β-cell secretory activity. By analogy with the euglycemic insulin clamp counterpart, the hyperglycemic clamp provides an M value, which represents the combined effect of endogenous hyperinsulinemia plus hyperglycemia on whole-body glucose disposal. The M value can be expressed per unit of insulin to provide an index of insulin sensitivity (M/I) that agrees closely with that derived from the euglycemic insulin clamp [62].

Intravenous glucose

In the presence of euglycemia, insulin displays a potent suppressive action on hepatic glucose production, such that portal insulin concentrations of less than 100 mU L⁻¹ abolish glucose entry into the circulation. Figure 14.7 shows a typical time course for endogenous glucose production following an acute increase in plasma insulin to levels of 60–70 mU L⁻¹ in a healthy subject. Dose–response curves relating calculated portal plasma insulin concentrations to suppression of glucose

Figure 14.7 Time course of suppression of hepatic glucose production in healthy adults during a euglycemic insulin clamp. Source: Redrawn from DeFronzo et al. 1983 [32].
production (Figure 14.8) indicate a half maximal effect at 30 mU L\(^{-1}\), corresponding to increments in portal insulin in the range of only 5–10 mU L\(^{-1}\) [63]. Note that in its capacity of a glucose-producing organ the liver is very sensitive to insulin; physiologic hyperinsulinemia, on the other hand, does activate glucokinase, but the resulting increment in hepatic glucose uptake is very small [64]. Hyperglycemia induced by intravenous glucose administration strongly synergizes this inhibitory action of insulin on hepatic glucose release (see Figure 14.4): in normal adults, a rise in arterial plasma glucose levels of only about 2 mmol L\(^{-1}\) is sufficient to reduce glucose output promptly by over 80% [28]. Figure 14.8 also shows the dose–response of insulin-stimulated whole-body glucose disposal. The apparent maximum at euglycemia is in the order of 60 μmol min\(^{-1}\) kg\(^{-1}\) in healthy adult subjects, whereas the half-maximum lies around 70–110 mU L\(^{-1}\) of peripheral (systemic) plasma insulin concentrations [32]. A dose–response curve of similar shape is derived when progressively higher insulin doses are applied locally in forearm tissues, about 70% of which consists of skeletal muscle [65]. Extrapolating the latter data to total body muscle mass it makes it possible to estimate that, with prevailing peripheral plasma insulin concentrations in the high physiologic range (60–90 mU L\(^{-1}\)), 50–70% of a total glucose flux of 30–40 μmol min\(^{-1}\) kg\(^{-1}\) is disposed of in muscle tissue. Obviously, this percentage increases further at still higher insulin levels as the contribution of insulin-independent tissues declines. The control of glucose production and utilization by insulin is dependent on both concentration and time. At any given hormone concentration, there is a finite time before the effect sets in and reaches its maximum. Such onset time is the sum of a circulatory delay (from arterial blood to cell surface) and a cellular lag (intracellular diffusion and effector activation). Similarly, insulin’s effect is present for some time (offset) after the circulating concentration has returned to prestimulatory levels. Figure 14.9 shows the activation and deactivation times of insulin calculated at euglycemia over a wide range of plasma hormone levels (up to 1000 mU L\(^{-1}\)) [66]. With the reservations inherent in the analysis of non-steady-state tracer data, these results provide evidence that activation and deactivation are inversely related to one another; thus, at higher insulin doses the effect is more rapid and takes longer to wane. Further, the relationship between onset and offset time is different for the liver (suppression of glucose release) and for peripheral tissues (stimulation of glucose uptake): at any insulin dose, the liver is activated more rapidly and more persistently. The latter phenomenon may have to do with the shorter diffusion time of blood-borne substances into highly perfused organs (1 mL min\(^{-1}\) per gram of tissue in the liver versus a corresponding value of 0.04 mL min\(^{-1}\) g\(^{-1}\) in resting skeletal muscle; Table 14.1). The inter-individual variation of insulin-stimulated glucose disposal is large, covering a five- to sixfold span even in relatively homogeneous groups of healthy subjects. Adipose mass and degree of physical fitness are powerful determinants of insulin sensitivity, in that weight loss and regular aerobic training are associated with demonstrable gains in insulin sensitivity. On the other hand, age, gender, distribution of body fat, diet, and menstrual phase are general physiologic covariates of insulin sensitivity. Evidence obtained in Pima Indians [67] demonstrates that genetic factors are at work in the distribution of insulin sensitivity in the population. By combining indirect calorimetry with dose–response studies using the clamp technique, it has been possible to quantitate the two major components of whole-body glucose disposal, that is, glucose oxidation and nonoxidative glucose disposal—the latter consisting of glycogen synthesis for the most part (>90%) and the remaining being net lactate production via aerobic glycolysis [68,69]. Figure 14.10 shows that the two daughter curves retain the sigmoidal shape of the mother curve but with distinctly different dose kinetics. Thus, glucose oxidation is more sensitive (lower apparent half-maximum) but saturates earlier (lower maximum) than glycogen synthesis; the latter behaves as a pathway with low sensitivity and high capacity.
Skeletal muscle is the predominant site of insulin-mediated net glycogen synthesis; insulin also has a potent effect to augment hepatic glycogen synthesis although its contribution to total body glycogen synthesis is small compared to skeletal muscle. The increment in carbohydrate oxidation that follows systemic insulin administration occurs in muscle as well as other tissues including liver and adipocytes. Insulin inhibits lipolysis, reduces plasma FFA levels, and decreases and lipid utilization very effectively [63]. As shown in Figure 14.11 (top), plasma FFA concentrations decline steeply in response to small increments in circulating insulin levels at euglycemia; this results from a drastic reduction in the rate of FFA appearance into the circulation. The consequence of the reduced availability of FFA is a parallel reduction in both FFA oxidation and nonoxidative FFA disposal, that is, reesterification (Figure 14.11, bottom). The reciprocal pattern of changes of glucose disposal/oxidation on the one hand, and lipid oxidation/utilization on the other, introduces the concept of substrate competition [70,71]. Glucose and long-chain FFA are the first and best-known example of substrates in mutual competition for use by insulin-dependent tissues as fuels. Physiologically, a rise in plasma glucose concentration increases the rate of glucose uptake into cells by mass action; the resulting increase in the generation of α-glycerolphosphate during anaerobic glycolysis supplies the substrate for augmented reesterification of tissue FFA, thereby limiting their release into the blood stream. This effect is reinforced by the glucose-induced rise in insulin, which further reduces the supply of lipid substrates by directly inhibiting lipolysis. This glucose-on-FFA feedback is balanced by an FFA-on-glucose negative feedback. A rise in FFA availability enhances lipid oxidation and restrains pyruvate oxidation (at the pyruvate dehydrogenase step). This sequence of biochemical events and the intracellular signals that trigger it were elucidated by Randle and coworkers in an elegant series of experiments in
Chapter 14

The glucose-FFA-amino acid cycle. Because glucose, FFA, and amino acids are all insulin secretagogues, isolated increases of each of them lower the circulating levels of the other two via hyperinsulinemia (inner ring). By substrate competition (outer ring), an increased supply of either FFA or amino acids will spare glucose. In addition, FFA have a hypoaminoacidemic effect on insulin-stimulated glucose metabolism by multiple other mechanisms. Following their entry into the cell, FFA are converted to their FAcO derivative. Long-chain FAcO inhibitors the insulin-signal transduction system by causing serine phosphorylation of the insulin receptor and insulin receptor substrate (IRS)-1 (see subsequent discussion) [72]. LC-FAcO also have independent effects to inhibit the glucose transport system (GLUT4), glucose phosphorylation (hexokinase), and glycogen synthase [42,74,76,77]. Elevated intracellular levels of LC-FAcO are associated with increased concentrations of diacylglycerol (DAG) and ceramides which also cause serine phosphorylation of the insulin receptor and IRS-1, resulting in impaired insulin signaling and insulin resistance [78–80]. Collectively, these negative effects of LC-FAcO, DAG, and ceramides on insulin-stimulated glucose metabolism have been referred to as lipotoxicity [81]. Amino acids also can enter a competition cycle with glucose, although somewhat less effectively than FFA. Increased amino acid provision enhances glucose production under conditions of insulin deficiency or resistance, and limits glucose utilization in the insulinized state [60]. Furthermore, raising FFA has a hypoaminoacidemic effect in humans [82].

In summary, each of the three major substrates, if present in excessive amounts (whether by endogenous production or exogenous administration), can lower the level of the other two by stimulating insulin release. In this capacity, glucose is obviously favored, being a much more potent secretagogue than fat or amino acids. In addition, multiple substrate effects (not mediated by changes in insulin release) participate in the regulation of the substrates themselves: high FFA and amino acids raise glucose, while high FFA lower amino acids [83] (Figure 14.13).

Figure 14.12 Dose-response curves for total glucose metabolism (a), glucose oxidation (b), and nonoxidative glucose disposal (c) in healthy subjects during euglycemic insulin clamps with (lipid) or without (control) a concomitant infusion of a triglyceride emulsion. Source: Thiebaut 1982 [124]. Reproduced with permission of Elsevier.

Figure 14.13 The glucose-FFA-amino acid cycle. Because glucose, FFA, and amino acids are all insulin secretagogues, isolated increases of each of them lower the circulating levels of the other two via hyperinsulinemia (inner ring). By substrate competition (outer ring), an increased supply of either FFA or amino acids will spare glucose. In addition, FFA have a hypoaminoacidemic effect of their own. See text for a detailed discussion.
Insulin actions *in vivo*: glucose metabolism

Cellular, biochemical, and molecular mechanisms of insulin action (see also Chapter 12)

**Insulin receptor/insulin receptor tyrosine kinase**

The insulin receptor is a glycoprotein consisting of two $\alpha$ subunits and two $\beta$ subunits linked by disulfide bonds and expresses insulin-stimulated kinase activity directed towards its own tyrosine residues [12,84–87]. Insulin receptor phosphorylation of the $\alpha$ subunit, with subsequent activation of insulin receptor tyrosine kinase, represents the first step in the action of insulin on glucose metabolism. Insulin receptors devoid of tyrosine activity are completely ineffective in mediating insulin stimulation of cellular metabolism. Mutagenesis of any of the three major phosphorylation sites (at residue 1158, 1163, and 1162) impairs the insulin receptor kinase activity, and this is associated with a marked decreased in the acute metabolic and growth-promoting effects of insulin [12,86–88].

**Insulin receptor signal transduction**

After activation, insulin receptor tyrosine kinase phosphorylates specific intracellular proteins, of which at least nine have been indentified (Figure 14.14). Six of these belong to the family of insulin-receptor substrate (IRS) proteins; IRS-1, IRS-2, IRS-3, IRS-4, IRS-5, IRS-6 (the others include Shc, Cb1, Gab-1, p60 [dok], and APS). In muscle, IRS-1 serves as the major docking protein that interacts with the insulin receptor tyrosine kinase and undergoes tyrosine phosphorylation in regions containing amino acid sequence motifs (YXXM or YMXM) that, when phosphorylated, serve as recognition sites for proteins containing src-homology 2 (SH2) domains. Mutation of these specific tyrosines severely impairs the ability of insulin to stimulate glycogen synthesis, establishing the important role of IRS-1 in insulin signal transduction [12,84–88]. In liver, IRS-2 serves as the primary docking protein that undergoes tyrosine phosphorylation and mediates the effect of insulin on hepatic glucose production, gluconeogenesis, and glycogen formation. In adipocytes, Cb1 is phosphorylated after its interaction with the insulin receptor tyrosine kinase and is required for stimulation of GLUT4 translocation [2].

In muscle, the phosphorylated tyrosine residues on IRS-1 mediate an association between the SH2 domain of the 85-kDa regulatory subunit of PI3K, leading to activation of the enzyme (Figure 14.14). PI3K is a heterodimeric enzyme composed of an 85-kDa regulatory subunit and a 110-kDa catalytic subunit. The latter catalyzes the 3'-phosphorylation of phosphatidylinositol (PI) in the plasma membrane glycolipids, thereby converting PI4,5-bisphosphate to PI3,4,5-triphosphate and PI4-phosphate to PI3,4 biphosphate. PI(3,4,5)P$_3$ and PI(3,4)P$_2$ lead to the stimulation of glucose transport [12,84–88]. Activation of PI3K by phosphorylated IRS-1 also stimulates glycogen synthase, via a process that involves activation of protein kinases B/Akt and subsequent inhibition of kinase such as GSK3 and activation of protein phosphatase-1 [89]. Insulin stimulates glycogen synthesis by simultaneously activating glycogen synthase and inhibiting glycogen phosphorylase [89,90] (Figure 14.15). The effect of insulin is mediated via the PI3K pathway, which inactivates phosphatases, particularly PP-1. It is believed that PP-1 is the primary regulator of glycogen metabolism [84,89–91]. In skeletal muscle, PP-1 associates with a specific glycogen binding regulatory subunit, causing dephosphorylation (activation) of glycogen synthase. PP-1 also phosphorylates (inactivates) glycogen phosphorylase. The precise steps that link insulin receptor tyrosine kinase/PI3K activation to stimulation of PP-1 have yet to be defined. Akt has been shown to phosphorylate and thus inactivate GSK3. This decreases glycogen synthase phosphorylation, leading to activation of the enzyme. From the physiologic standpoint, it makes sense that activation of glucose synthesis

**Figure 14.14** Schematic representation of the insulin receptor and the cascade of intracellular signaling molecules that have been implicated in insulin action. See text for a more detailed discussion.
transport and glycogen synthase should be linked to the same signaling mechanisms to provide a coordinated and efficient stimulation of intracellular glucose metabolism.

**Glucose transport**

Activation of the insulin signal transduction system in insulin target tissues leads to the stimulation of glucose transport. The effect of insulin is brought about by the translocation of a large intracellular pool of GLUTs (associated with low density microsomes) to the plasma membrane [1,2,84,92]. As discussed earlier, there are five major, different facilitative GLUTs with distinctive tissue distributions. GLUT4, the insulin regulatable transporter, is found in insulin-sensitive tissues (muscle and adipocytes), has a $K_m$ of $\sim 5 \text{ mmol L}^{-1}$, which is close to that of the fasting plasma glucose concentration and is associated with HK-II [19,93]. In adipocytes and muscle, its concentration in the plasma membrane increases markedly after exposure to insulin, and this increase is associated with a reciprocal decline in the intracellular GLUT4 pool. Acute physiologic hyperinsulinemia does not increase the total number of GLUT4 in muscle, even though several studies have demonstrated an increase in muscle GLUT4 mRNA. Using a novel isotopic dilution technique, the in vivo dose–response curve for the action of insulin on glucose transport in human forearm skeletal muscle has been described [94,95] (Figure 14.16). GLUT1 represents the predominant GLUT in the insulin-independent tissue (brain and erythrocytes). It is located primarily in the plasma membrane, where its concentration changes little after the addition of insulin. It has a low $K_m$ ($\sim 1 \text{ mmol L}^{-1}$) and is well suited for its function, which is to mediate basal glucose uptake. It is found in association with HK-I [19,93,96]. GLUT2 predominates in the liver and pancreatic $\beta$ cells, where it is found in association with a specific HK, HK-IV. In the $\beta$ cell, HK-IV is referred to as glucokinase [97]. GLUT2 has a high $K_m$ ($\sim 15–20 \text{ mmol L}^{-1}$) and, as a consequence, the glucose concentration in cells expressing this transporter increases in direct proportion to the increase in plasma glucose concentration. This characteristic allows these cells to behave as glucose sensors.

**Glucose phosphorylation (Table 14.2)**

Glucose phosphorylation and glucose transport are tightly coupled phenomena. Isoenzymes of HK (HK-I to HK-IV) catalyze the first committed intracellular step of glucose metabolism, the conversion of glucose to G-6-P [19,93,96,97]. HK-I, HK-II, HK-III are single-chain peptides that have a number of properties in common, including a very high affinity for glucose and product inhibition by G-6-P. Glucokinase (HK-IVB) is believed to be the glucose sensor in the $\beta$ cell, whereas HK-IVL plays an important role in the regulation of hepatic glucose metabolism.

In human skeletal muscle HK-II transcription is regulated by insulin [98]. HK-I is also present in human skeletal muscle but is not regulated by insulin. In response to physiologic euglycemic
hyperinsulinemia, HK-II cytosolic activity, protein content, and mRNA levels increase by 50% to 200% in healthy subjects [98] and this is associated with the translocation of HK-II from the cytosol to the mitochondria [99]. In contrast, insulin has no effect on HK-I activity, protein content, or mRNA levels.

**Glycogen synthesis (Figure 14.15)**

After glucose enters the cell or is phosphorylated, it can either be converted to glycogen or enter the glycolytic pathway. Of the glucose that enters the glycolytic pathway, approximately 90% is oxidized. In the low physiologic range of hyperinsulinemia, glycogen synthesis and glucose oxidation are of approximately equal quantitative importance. With increasing plasma insulin concentrations, glycogen synthesis becomes predominant [63,68]. Glycogen synthase is the key insulin-regulated enzyme that controls the rate of muscle glycogen formation [100–102]. Insulin enhances glycogen synthase activity by stimulating a cascade of phosphorylation-dephosphorylation reactions that ultimately lead to the activation of PP-1 (also called glycogen synthase phosphatase). The regulatory subunit (G) of PP-1 has two serine phosphorylation sites, called site 1 and site 2. Phosphorylation of site 2 by cyclic adenosine monophosphate-dependent kinase (PKA) inactivates PP-1, while phosphorylation of site 1 by insulin activates PP-1, leading to the stimulation of glycogen synthase. Phosphorylation of site 1 of PP-1 by insulin in muscle is catalyzed by insulin-stimulated protein kinase-1, which is part of a family of serine/threonine protein kinases termed ribosomal S6-kinases. The effect of insulin on glycogen synthase gene transcription and translation in vivo has been studied by employing the Glucose sensor classification of glucose transporter and hexokinases in various tissues.

<table>
<thead>
<tr>
<th>Organ</th>
<th>Glucose transporter</th>
<th>Hexokinase Coupler</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brain</td>
<td>GLUT1</td>
<td>HK I</td>
<td>Glucose dependent</td>
</tr>
<tr>
<td>Erythrocyte</td>
<td>GLUT1</td>
<td>HK I</td>
<td>Glucose dependent</td>
</tr>
<tr>
<td>Adipocyte</td>
<td>GLUT4</td>
<td>HK II</td>
<td>Insulin dependent</td>
</tr>
<tr>
<td>Muscle</td>
<td>GLUT4</td>
<td>HK II</td>
<td>Insulin dependent</td>
</tr>
<tr>
<td>Liver</td>
<td>GLUT2</td>
<td>HK IVa</td>
<td>Glucose sensor</td>
</tr>
<tr>
<td>B-cell</td>
<td>GLUT2</td>
<td>HK IVb</td>
<td>Glucose sensor</td>
</tr>
<tr>
<td>Gut</td>
<td>GLUT3-Symporter</td>
<td>Sodium dependent</td>
<td></td>
</tr>
<tr>
<td>Kidney</td>
<td>GLUT3-Symporter</td>
<td>Sodium dependent</td>
<td></td>
</tr>
</tbody>
</table>

**Glycolysis/glucose oxidation**

Glycolysis/glucose oxidation accounts for approximately 90% of total glycolytic flux, while anaerobic glycolysis accounts for the other 10% [69]. Two enzymes, PFK and PDH, play central roles in the regulation of glycolysis and glucose oxidation, respectively. PFK represents a key functional step in control of glycolysis [103,104]. However, insulin does not exert any direct effect on this enzyme, which is primarily regulated by the energy (ATP) and fuel (citrate, acetyl-CoA) status of the cell. However, insulin indirectly stimulates PFK by increasing fructose-2,3-bisphosphate, a potent activator of PFK. Insulin has no effect on muscle PFK activity, mRNA levels, or protein content in nondiabetic individuals [105]. Insulin also regulates flux through glycolysis by increasing the activity of the multienzyme complex, PDH [106,107]. This enzyme is activated by insulin, which stimulates PDH phosphatases, thus converting the enzyme from its inactive phosphorylated form to its active dephosphorylated form (Figure 14.17). The PDH complex enzyme is also inhibited by its products, acetyl-CoA and reduced nicotinamide adenine dinucleotide (NADH).

**Oral glucose**

At any point in time, the glycemic response to exogenous glucose is the balance between the rate at which glucose appears in the systemic circulation (from oral as well as endogenous sources) and the rate at which glucose is disposed of. Oral glucose appearance in the peripheral circulation depends on: (a) the rate at which the gastric contents are passed on to the small intestine; (b) the rate of intestinal glucose absorption; (c) the extent of gut glucose utilization; (d) the degree of hepatic glucose trapping; and (e) the dynamics of glucose transfer through gut, liver, and posthepatic circulation on to the right heart. The contribution of endogenous glucose to the glycemic response to feeding depends on the extent and rate of change of hepatic glucose production. Finally, glucose disposal depends on changes in the pattern of hormonal stimuli and substrate availability.

Being a summation phenomenon, the response to oral glucose explores the whole of glucose tolerance, not the individual contribution of the various components. The rate-limiting step in the transfer of ingested glucose from the stomach to the liver is...
the rate of gastric emptying. This depends on the volume, temperature, and osmolarity of the glucose solution in the case in which glucose alone is ingested. Glucose absorption through the intestinal lining cells is rapid and efficient, the capacity of the whole small intestine being far in excess of ordinary needs. The presence of sodium chloride in the glucose drink enhances glucose absorption. Glucose utilization by intestinal tissues is small when glucose is presented on the vascular side, that is, when there is no oral glucose (Table 14.1).

The possibility also exists that systemic hyperglycemia and/or hyperinsulinemia may interfere with intestinal glucose absorption. Experimentally induced hyperglycemia impairs gastric motility and slows down gastric emptying, thereby delaying the absorption of oral antidiabetic drugs.

Glucose uptake by the liver is stimulated by portal hyperglycemia (see later), but traversing the hepatic space is unlikely to significantly delay the appearance of oral glucose in the systemic circulation. On the whole, the dynamics of oral glucose appearance are essentially driven by gastric emptying, while intestinal transit, crossing of the mucosa to enter portal blood, and transhepatic passage together introduce only a small time delay. In other words, if neither gut nor liver tissues used glucose, the time course of oral glucose appearance would only follow gastric emptying, with a time shift of a few minutes. For this reason, the absorption step is a major component of the shape of the glycemic response to glucose. Figure 14.18 shows the pattern of appearance of ingested glucose in the systemic circulation in healthy individuals, as reconstructed by a double tracer technique [108]. Glucose arrival peaks within 30–45 min, declines slowly thereafter, and is still significantly above zero 210 min after glucose ingestion. A secondary rise in oral glucose appearance is sometimes seen between 2 and 3 h after ingestion. Figure 14.18 also shows the time course of suppression of endogenous glucose release by oral glucose. A sustained nadir between 45 and 135 min is followed by a slow return toward fasting rates; hepatic glucose production is still significantly inhibited 210 min after the glucose challenge. Overall suppression of endogenous glucose production during 3–4 h after ingestion averages 50%, surprisingly less than what would be expected on the basis of the combined portal hyperglycemia and hyperinsulinemia (see Figure 14.4). Relative to intravenous glucose/insulin administration, glucose ingestion evidently reinforces counterregulatory influences which keep liver glucose outflow open.

In Figure 14.19 the observed arterial plasma glucose concentration is broken down into the component contributed by oral glucose appearance and that provided by hepatic glucose production. While the resemblance of oral Ra to the plasma glucose curve is evident (especially during the first 60–90 min), less appreciated is the fact that absorption is still incomplete 3–4 h after ingestion. Figure 14.18 depicts the time course of total glucose disposal (Rd) following oral glucose: with a lag of some 30 min, glucose uptake is stimulated by 50–110% throughout the period of observation. Hyperglycemia contributes more to whole-body glucose disposal during the first half of the test; thereafter, hyperinsulinemia predominates. Oral glucose elicits vasodilation of the splanchnic vascular bed; this, too, is a change that persists for at least 4 h [109]. Thus, both the metabolic and the hemodynamic perturbations induced by oral glucose extend beyond the time of return of plasma glucose to pre-ingestion levels. The tissue destination of absorbed glucose has been the subject of intense investigation. While the liver classically was reputed to be responsible for the eventual disposal of the majority of oral glucose [109], the weight of more recent evidence [33,36,110] favors the view that peripheral tissues are responsible for between one half and two thirds of glucose uptake, while the splanchnic tissues account for the remainder.

A robust insulin secretory response directs more posthepatic glucose to the periphery, while a large increase in splanchnic blood flow increases the delivery of incoming sugar to the liver. In humans, for example, a glucose drink sipped over 3.5 h rather than swallowed in one bolus generates the same overall glucose curve but a 50% smaller endogenous insulin response [111].

The route of administration seems to influence the metabolic fate of glucose [33,36,113,114], in that the portosystemic glucose gradient per se enhances liver glucose uptake independently of portal glycemia and total glucose delivery to the organ [33,36,112,113] (as previously discussed).
Limited information is available on the intracellular fate of ingested glucose. While glucose oxidation in the brain continues unabated during the absorptive period, some 50% of the glucose taken up by peripheral tissues (muscle) is oxidized, the remainder being stored as muscle glycogen or as lactate in the lactate pool [114]. During absorption, there is an increase in lactate release by both the splanchnic area as a whole and the intestine [28–30]. In the latter, it has been estimated that some 5% of the ingested load is converted into three-carbon precursors of glucose (lactate, pyruvate, and alanine) and passed on to the liver [28,115]. The net release of lactate by the splanchnic area indicates that the sum of hepatic lactate production and gut lactate formation exceeds hepatic lactate extraction. Liver glycogen formation during absorption of oral glucose certainly occurs both directly from glucose and indirectly via gluconeogenesis. The relative contribution of the direct versus indirect pathway to hepatic glycogen synthesis is somewhat uncertain owing to methodological difficulties. Current data [116] suggest that gluconeogenesis participates in liver glycogen repletion to a lesser extent in humans than in rats [117,118].

Following glucose ingestion, the plasma insulin response is two- to threefold greater than that observed when the same glucose profile is created by intravenous glucose [119]. This incretin effect is due to the release of glucagon-like peptide (GLP)-1 and glucose-dependent insulinotropic polypeptide (GIP), which are secreted by the K cells of the early small intestine and L cells of the distal small bowel/large intestine [119,120]. GLP-1 and GIP are released within 2–3 min following glucose ingestion. Since gastric emptying takes 10–15 min to begin, it is clear that the nutrients cannot have reached the duodenum and certainly not the large bowel. The release of GLP-1 and GIP is mediated via neural connections to the brain and back via the vagus nerve [121]. Although activation of the neural pathway within minutes leads to the release of GLP-1 and GIP, insulin secretion is not increased since the stimulatory effect of GLP-1 and GIP on the β cell is glucose dependent and only occurs when gastric emptying begins and the plasma glucose concentration rises. Approximately half of all insulin that is secreted in response to oral glucose is dependent upon GLP-1/GIP in healthy subjects. A direct effect of GLP-1 to augment hepatic glucose uptake, beyond its stimulatory effect on insulin secretion, has been proposed [122]. The inhibition of glucagon secretion and simultaneous stimulation of insulin secretion conspire to suppress hepatic glucose production. GLP-1 also has a direct effect on the liver to enhance hepatic glucose uptake [123].

Key points
- The liver (hepatic glucose production) is more sensitive to the inhibitory action of insulin than are peripheral tissues (glucose uptake) to the stimulatory action of the hormone.
- Insulin action on glucose metabolism is both direct (enhancement of glucose transport, glycolytic breakdown, and incorporation into glycogen) and indirect (inhibition of lipolysis, lipid oxidation, and protein degradation).
- In insulin-sensitive tissues, the three major substrates (glucose, FFA, amino acids) are in competition with one another.
- The counterregulatory hormones oppose both the direct and the indirect actions of insulin on glucose disposal.
- Physiologic hyperinsulinemia activates the insulin signal transduction system leading to a coordinated increase in glucose transport/phosphorylation, glycogen synthesis, and glucose oxidation.
- The shape of the oral glucose tolerance curve is dominated by absorption processes, especially early after ingestion. Later on, insulin action prevails.
- Insulin controls oral glucose tolerance by suppressing endogenous glucose production and promoting glucose uptake into muscle. Hyperglycemia enhances splanchnic glucose uptake.
- GLP-1 directly, and indirectly by increasing insulin and inhibiting glucagon, augments hepatic glucose uptake and inhibits hepatic glucose production.

References


74 Pendergrass M, Nucci G, DeFronzo RA: In vivo glucose transport (GT) and phosphorylation (GP) in skeletal muscle are impaired by elevation of plasma FFA. Diabetes 1998;47:A64.
Insulin actions *in vivo*: glucose metabolism


CHAPTER 15

Measuring insulin action in vivo

Richard N. Bergman and Marilyn Ader
Cedars-Sinai, Los Angeles, CA, USA

Key points

• Insulin resistance is a risk factor for many diseases including diabetes.
• The glucose clamp remains the most straightforward and accurate method, but it must be done carefully.
• The minimal model method is accurate, and also yields indices of insulin response and metabolic clearance rate of insulin, and glucokinase activity.
• Surrogate measures of insulin resistance, HOMA and QUICKI, are not accurate with even a moderate depression of $\beta$-cell function.
• OGTT measures can be used, but are also not independent of changes in glucose absorption, insulin secretion, and insulin clearance.
• There is a need for more accurate noninvasive measures of insulin action.

Historical perspective

The heralded isolation of insulin in Toronto in 1921 was followed immediately by treatment of diabetes. It soon became clear that while insulin was effective in regulating the blood glucose levels in most patients, there were some subjects in whom insulin appeared to be ineffective [1,2]. This lack of insulin effect was termed “insulin resistance” as early as 1925 [3]. In a remarkable early study, Bainbridge observed that rats and mice, fed a carbohydrate-free, excess fat diet, developed a high degree of resistance to insulin. In 1936, Himsworth noted that the effect of insulin injection was less pronounced in obese subjects and subjects with “mild” diabetes [4]. He suggested that obesity and diabetes are related to insulin resistance [5] and that insulin sensitivity could be increased by high carbohydrate, low fat diets [6]. However, lack of ability to measure insulin directly made the interpretation of Himsworth’s elegant studies difficult in terms of insulin action per se. Introduction of the radioimmunoassay by Berson and Yalow provided the first compelling evidence for the existence of insulin resistance. These authors were aware that higher insulin after oral glucose implied insulin resistance, and suggested insulin resistance in obese patients, and patients with “maturity onset” (type 2) diabetes [7,8]. However, absent a direct assessment of insulin resistance, the putative role of this presumed defect in type 2 diabetes (T2DM) remained controversial. Adding to the controversy was Kipnis’ elegant demonstration of an insulin secretory defect in type 2 patients. Kipnis reported a diminished plasma insulin response to a matched plasma glucose pattern in type 2 diabetic patients compared to nondiabetic patients [9]. Thus as of 1970, supportive evidence existed suggesting a defect in $\beta$-cell function was responsible for “maturity onset” diabetes; au contraire, insulin resistance could likewise be implicated.

Opening the loop: the pancreatic suppression test

Controversy regarding insulin resistance in T2DM was due to the “closed-loop” feedback system, which regulates the blood glucose concentration (Figure 15.1). In normal individuals, ingestion of a meal increases glucose and other secretagogues, including GIP and GLP-1, which act in concert to stimulate insulin secretion. Hyperinsulinemia in turn acts to renormalize the plasma glucose level by suppressing glucose production and increasing glucose utilization. In the presence of insulin resistance per se, the elevated insulin will be less able to normalize glucose, therefore resulting in a secondary stimulus to the $\beta$ cells and relative hyperinsulinemia. In fact, postprandial hyperinsulinemia was considered by Berson and Yalow to be the signature for insulin resistance [7,8]. However, $\beta$-cell oversecretion could also account for postprandial hyperinsulinism. In fact, some investigators have suggested that hyperinsulinism itself is the primary event in the development of T2DM [10]. Because of the closed-loop relationship between insulin secretion and insulin action it is problematic to infer the existence of insulin resistance directly from a “closed-loop” procedure such as the oral glucose
tolerance test (OGTT), although this approach has become more widely used lately (see later, [11,12]). Equally important, it appeared it was not possible to assign a quantitative value to insulin resistance based solely on the closed-loop response to oral glucose.

Shen and Reaven demonstrated definitively the existence of insulin resistance in patients with T2DM [13–15]. These investigators intervened experimentally to interrupt the closed-loop. They made the β cells relatively unresponsive to increasing ambient blood glucose. Initially this was done with epinephrine [14,15] and later with somatostatin [16,17]. Infusing a compound that renders the β cells unresponsive during glucose infusion, the resultant steady-state glucose concentration “SSPG” is suggestive of insulin resistance. They reported that SSPG was higher in obesity, but highest in T2DM [18,19]. Thus, Reaven and colleagues provided persuasive evidence that insulin resistance existed concomitant with T2DM. However, the existence of insulin resistance did not rule out the possibility that β-cell deficiency could also play a role in the pathogenesis of T2DM. Also, SSPG, while indicative of apparent insulin resistance, does not represent a quantitative in vivo index of insulin sensitivity of the tissues. In addition, SSPG is highly dependent on utilization of glucose at high concentrations independent of the ambient insulin level (“glucose effectiveness,” cf. [20]). In fact, the debate regarding the relative importance of resistance versus insulin secretory defects in T2DM raged for several decades beyond the introduction of the pancreatic suppression test, which confirmed the insulin resistant state [21–25].

**Qualitative versus quantitative**

Often it is not sufficient to establish whether insulin resistance exists, relative to a normative population. Much more common is the need to measure insulin resistance, that is, to attach a number to it. A drawback of the qualitative resistance concept is that (as discussed earlier) virtually complete insensitivity to insulin can exist (“infinite insulin resistance”). It is much more practical to think in terms of “insulin sensitivity,” where the latter concept is the inverse of resistance. Thus, insulin sensitivity determined by whatever method may have a minimum value near zero (total resistance), and values may extend to a finite number representing great sensitivity to the hormone. The first quantitative method to measure insulin sensitivity was originated by the late Ruben Andres, and it is still called the “euglycemic glucose clamp” [26].

**The glucose clamp**

Rather than using exogenous agents that suppress β-cell insulin secretion, the glucose clamp uses external feedback control to “open the loop” between insulin secretion and sensitivity (Figure 15.2). The glucose clamp is a powerful and widely used method to attain a quantitative measure of insulin sensitivity, and has been applied in many hundreds of experimental studies.

**Performing the clamp**

DeFronzo and colleagues published the most widely applied clamp methodology [26,27]. In the simplest “euglycemic” manifestation, catheters are inserted in the right and left arms of the subject, and basal samples are collected, after which insulin is infused intravenously at a constant rate. The dose of insulin has varied widely [26–34]; however, the usual dose is 40 mU min⁻¹ m⁻². Infusing insulin would normally lower the blood glucose; however, the lowering is prevented by infusing glucose concomitant with the insulin infusion. It is not possible a priori to know the rate of glucose infusion to prevent insulin from changing; therefore, a separate external feedback loop is established wherein glucose is measured frequently (usually every 5 min) and infusion rate is chosen to attempt to keep the plasma glucose concentration from changing.

In normal subjects plasma glucose concentration is usually “clamped” at or near 5 mM. Clearly, if a subject is extremely insulin resistant (sensitivity ~ 0), very little exogenous glucose would be needed to be infused despite increased insulin; if a...
subject is very sensitive to insulin, the needed exogenous glucose infusion rate will be substantial. The rate of glucose infusion is considered a reflection of insulin sensitivity. Insulin sensitivity is normalized to body size in some manner—either per kg body weight [35], per surface area [36,37], or per unit fat-free mass [38,39].

It is not a simple matter to determine in real time the exact glucose infusion rate needed to keep glucose at a predetermined “target” value. There have been numerous attempts to automate the external feedback loop, which includes the glucose measuring device as well as the glucose infusion device. Algorithms have been made available to calculate glucose infusion rate independently [26,40–42]. These algorithms have the advantage of objectivity; however, in most laboratories even today exogenous glucose infusion rate is determined at the bedside by a trained individual. The need for frequent monitoring of the blood glucose, and the need for a practiced individual to control the glucose infusion rate as a function of time, renders the glucose clamp method labor intensive and costly. Although providing an accurate measure of insulin sensitivity, under most circumstances three trained individuals are needed to carry on a safe and effective glucose clamp experiment on a human subject.

**Components of insulin action calculated from the euglycemic clamp**

Under fasting conditions in normal individuals, the blood glucose is determined by a balance between endogenous glucose production by the liver (and kidneys) and glucose uptake by insulin-independent (brain, red blood cells, gut) and insulin-dependent tissues (skeletal muscle, heart, adipose). Infusing insulin lowers the blood glucose both by inhibiting endogenous glucose production (EGP) and by stimulating glucose uptake (Rd). One of the great assets of the glucose clamp is the potential for teasing apart these two effects of insulin on glucose handling. Over three decades accurate methods were established to measure rates of endogenous glucose production—even under conditions of substantial glucose turnover that is observed during clamps [26,43–46]. These methods involve the infusion of glucose either labeled with tracer (usually 3-3H-glucose), or stable isotopes such as 6,6 dideuterated glucose. Glucose output is calculated based upon the dilution by unlabeled endogenously produced glucose of the labeled glucose pool in extracellular fluid (Figure 15.3).

Great care is necessary to make accurate calculations of endogenous glucose production and glucose uptake during glucose clamps [26,45–50]. The large potential error is due to the considerable glucose turnover stimulated by hyperinsulinemia. Endogenous production is the difference between two large numbers: total rate of glucose disposal minus the rate of exogenous glucose infusion. Small errors in estimating either of these rates will cause a large error in calculation of glucose output. Finegood and colleagues proved that to make an accurate calculation, exogenous glucose must be labeled at a specific activity or enrichment similar to that attained when labeled moiety is infused into the fasting patient [44]. Under such conditions plasma specific activity remains near steady-state and accurate EGP is calculated. Impurities in commercial labeled glucose preparations can lead to spurious results [49]; thus tracer should be checked and purified if necessary. Also, even being unaware that so-called 50% D/W dextrose solution is only 45.4% glucose can lead to large errors. Thus, it is important to be vigilant in these calculations and to be familiar with the literature on these subjects.

Based upon accurate methods applied to normal individuals, suppression of glucose output by liver is much more sensitive to insulin ($ED_{50} \sim 25 \text{mU L}^{-1}$) than stimulation of glucose uptake ($ED_{50} \sim 60 \text{mU L}^{-1}$) [48,50]. Reduced insulin sensitivity as assessed by the clamp method has been demonstrated in an enormous variety of physiologic and pathophysiologic situations including fasting, impaired glucose tolerance (IGT), polycystic ovarian syndrome, pregnancy, children of low birth weight, sedentary lifestyle, obesity, and acromegaly, as well as many ethnic groups from around the globe. Comparing insulin resistance in these different conditions and populations is important because insulin resistance is a primary risk factor for diabetes, but also for many other chronic diseases, including cardiovascular disease, hypertension, and cancer.

**A quantitative index from the glucose clamp**

As stated, glucose infusion rate during the clamp is an index of insulin sensitivity. Because insulin suppresses endogenous glucose output and increases glucose disposal, infusion of glucose must compensate for both changes under hyperinsulinemic but euglycemic conditions; the infusion rate is thus the sum of these two effects:

$$\text{Glucose infusion rate} = \text{Increase in glucose disposal} + \text{Suppression of endogenous glucose production}$$
Changes in either of these components on the right-hand side of this equation may be regarded as contributing to insulin sensitivity; percent suppression of EGP is often considered hepatic insulin sensitivity, whereas increase in glucose disposal (Rd: expressed as mass per unit time per unit body size) is reflective of “peripheral” insulin sensitivity (usually considered the effect of hyperinsulinemia on glucose uptake by skeletal muscle). In populations of normal subjects knowing EGP and Rd changes are often sufficient; however, in pathologic conditions they may be misleading. Glucose uptake is a nonlinear function of glucose and insulin [43,51,52]. It therefore may not be correct to compare Rd values in individuals assessed at different glucose levels or at different insulin levels [43,51]. Generally subjects with fasting hyperglycemia may have their fasting glucose values normalized by infusing exogenous insulin acutely, or overnight [53–55].

Insulin is degraded very rapidly under normal conditions. Because the liver degrades 50% or more of the insulin presented to it during a single passage of blood from the abdominal portal vein to the hepatic veins, the plasma level of insulin achieved during a euglycemic clamp can vary substantially. Generally insulin clearance is reduced in insulin-resistant situations, resulting in proportionately greater plasma insulin levels during insulin infusion during the clamp [56–58]. For example, in the canine model, simple fat feeding for 6 weeks decreased fractional insulin clearance by the liver from 60 to 40% for a single passage [59]. To compensate for differences in insulin clearance, many investigators divide glucose disposal (Rd) by the increment in insulin under steady-state conditions (M/I).

We proposed a single variable, the “insulin sensitivity index,” which attempts to correct or normalize for differences in plasma glucose and insulin concentration at which clamps may be performed [60–63]. The clamp-based index is defined as the following:

\[
\text{SIP (Clamp)} = \frac{\Delta \text{Rd}}{\Delta \text{Insulin}} \quad \times \text{Glucose}
\]

in which SIP represents peripheral insulin sensitivity, the numerator is the increase in glucose disposal during a clamp, and the denominator is the product of the insulin increment during the clamp, and the ambient glucose at which the clamp is performed. Thus, this equation calculates increase in glucose clearance per unit increase in the plasma insulin value (Figure 15.4).

**Insulin action dynamics**

Insulin sensitivity from the clamp is expressed in terms of the steady-state value reached after a period of hyperinsulinemia. A dose–response can be constructed and sensitivity can be expressed as the ED50, that is, the concentration for half-maximal stimulation of Rd as a function of dose. Alternatively, sensitivity can be expressed (as it usually is) as the Rd at a specific dose of insulin, which may be low, intermediate, or a maximum dose. It is often assumed that steady-state rate of glucose uptake is reached by 180 min after onset of the insulin infusion. In fact, steady-state glucose uptake is not reached at 3 h [64,65]. Glucose uptake rate at 3 h is only 2/3 the “true” steady-state, which is not achieved until 6 h (Figure 15.5). Certainly, shorter periods (e.g. 120 min) are inadequate to reflect steady-state rates of uptake [43,63,66]. Rd at “steady-state” does not incorporate the events that account for the slow time course of activation of glucose uptake [43,63,67]. Events that may determine the time course of glucose uptake in the face of hyperinsulinemia include dynamics of plasma insulin itself [66], access of insulin to the insulin-sensitive tissues (primarily skeletal muscle, liver, and adipose tissue) [68,69], binding of insulin to receptors, and the downstream signaling pathways and mobilization of glucose transporters (primarily glucose transporter 4 (GLUT4; [70–77])). It has been of interest to examine the in vivo pathway of insulin action during clamps to identify the rate-limiting step which determines insulin dynamics in vivo. It is possible
that changes in access to insulin-sensitive cells could contribute to a “dynamic” insulin resistance, because insulin action does not generally reach the hyperinsulinemic steady-state value assessed during euglycemic clamp studies.

Access to sensitive cells includes the actual process of movement of insulin across the endothelial cells as well as any effect of insulin to access so-called “nutritive” capillaries (i.e., those that perfuse insulin-sensitive and glucose-utilizing or glucose-storing tissues). By measuring insulin in plasma and lymph simultaneously, we were able to determine that it was the movement of insulin between the plasma and the interstitial fluid compartment that was the determinant of insulin action dynamics [68,78].

It is clear that cellular sensitivity to the hormone is not the only action of insulin on carbohydrate metabolism which results in insulin resistance. The ability of the hormone to access sensitive tissues is reduced because of reduced transendothelial transport as well as failure to increase flow through nutritive capillaries. The relative importance of these dynamic components of insulin action under different conditions remains to be established. Given the various dynamic features that contribute to overall insulin action, it is reasonable to utilize the dynamics of insulin action, rather than quasi steady-state alone to represent in vivo insulin sensitivity. One approach that utilizes insulin action dynamics is the “minimal model” approach.

**Dynamic insulin action: the minimal model**

In the normal 24-h day, the insulin-sensitive tissues are never exposed to steady-state conditions. After a meal, glucose and insulin are changing, and the rate of glucose uptake lags in time behind the time courses of glucose and insulin. It may be preferable to measure insulin action from a dynamic relationship between glucose, the primary nutritive carbohydrate, insulin, the primary anabolic hormone, and the rate of glucose disposition. One approach to such a measurement is the frequently sampled intravenous glucose tolerance test (FSIGT) coupled with the “minimal model.” The underlying concept behind the minimal model approach will be outlined herein, but the reader should refer to more extensive presentations for added information and detail [43,60,79–83].

**Concept**

As discussed, intervening to “open the loop” to perform a glucose clamp requires resources in dollars and labor. Using the power of the digital computer and modeling obviates the need to maintain constant plasma glucose and insulin to obtain a sensitivity measure. To measure dynamic sensitivity the insulin/glucose regulating system must be perturbed. While the administration of oral glucose might be considered the simplest approach given the acceptance of the OGGT as a diagnostic procedure, oral glucose has substantial limitations due to lack of control over the rate of glucose appearance in the extracellular fluid [84,85] (see later). Therefore, intravenous administration was the protocol of choice. The explicit detailed dynamic relationship between glucose and insulin following intravenous glucose is measured by collecting frequent blood samples (Figure 15.6).

While many hormonal and neural signals can impact plasma dynamics, it has been shown that in the main, a few factors dominate the time course of glucose and insulin following glucose injection. These are (1) endogenous glucose production (by liver predominantly and also kidney); (2) insulin secretion and insulin clearance; (3) effect of insulin to alter the rate of glucose uptake by insulin-dependent tissues (mostly skeletal muscle) by a variety of mechanisms including transporter mobilization, enzyme activation, changes in perfusion of nutritive tissues, endothelial transport; and (4) glucose uptake by tissues independent of insulin (brain, gut, red blood cells).

**The model**

It was a challenge to develop a computer model that could account for plasma dynamics in a compact but accurate package. The model was the simplest representation we could devise, which was based upon known physiology, and which could accurately describe moment-by-moment plasma dynamics. The model continues to thrive in that it continues to be the basis for a large number of clinical investigations (~50 per year) as well as a robust literature related to its mathematical and computer characteristics [43,60,61,79–82,86–96].

**Practical application of the minimal model**

In contrast to the glucose clamp procedure, neither online measurements nor external control of infusions are necessary to perform the minimal model procedure. Temporary intravenous lines are placed in antecubital veins in a relaxed...
subject. After basal sampling, glucose is injected and frequent sampling is begun. To facilitate the computer’s ability to estimate metabolic parameters (cf. Table 15.1) insulin is injected at 20 min and sampling continued until 3 h [92,93]. Samples are spun and plasma is stored frozen for later analysis of glucose and insulin. It is straightforward for two experimenters to perform the FSIGT on one subject.

Yields from the minimal model

By computer analysis using the program MINMOD Millenium [97], a set of values is calculated for each individual subjected to the FSIGT test (Table 15.1). Insulin sensitivity from the model is expressed as parameter $S_t$. The latter parameter from the minimal model is fundamentally analogous to the previously mentioned $S_{IP(clamp)}$ estimated from the euglycemic clamp [60]. In fact, the parameters are strongly correlated with each other [60–62,98–100]. The correlation is not perfect, however. This may be due to inherent errors in measurement of each (errors in plasma measurements, sampling times, etc.). Also, it may be due to the fact that the minimal model $S_t$ is estimated from a dynamic test, whereas the analogous clamp parameter is from a steady-state test [82]. It is apparent that these very different tests—the FSIGT versus the clamp—may not be measuring the exact same things, but what they are reflecting are highly correlated. The clamp emphasizes steady-state effects of insulin on glucose uptake and production often at a single insulin level. The minimal model also measures the factors that determine the rates of distribution of insulin as well as glucose among different compartments of the organism, and factors that determine how rapidly insulin can stimulate glucose metabolism by accessing tissues (whether by changing blood flow distribution or endothelial transport) and by mobilizing transporters, activating enzymes, changing expression, and so on. The relative importance of insulin resistance to pathogenesis disease, whether measured by a steady-state or a dynamic approach, is of great interest.

Additional important information is gleaned by using the minimal model approach. One important parameter, also obtained from the euglycemic clamp (by the ratio of the increment in plasma insulin to the insulin infusion rate) is an estimate of the metabolic clearance rate of the hormone. As shown in Figure 15.6, insulin is injected into the patient 20 min after the glucose. The rate of decline in the insulin level after this injection is proportional to the rate of insulin clearance. The metabolic clearance rate of insulin is estimated by the following equation:

$$MCRI(l/kg/min) = \frac{Dose}{\int_{t=20}^{\infty} (Ins(t) - Ins(0))dt}$$

where $MCRI$ represents the metabolic clearance rate of insulin, $Dose$ reflects the amount of insulin injected at $t = 20$ min during the FSIGT (in mU kg$^{-1}$), $Ins(t)$ is the plasma insulin concentration at time $t$, and $Ins(0)$ is the insulin concentration at $t = 0$ (=basal insulin, in mU L$^{-1}$).

However, this estimate is imperfect, because the time course of plasma insulin is also modified based upon the change in the endogenous release of the hormone. In our laboratory, we are developing new methods for providing a more accurate estimate of insulin clearance, which utilizes the C-peptide measurements made following the insulin injection at 20 min after glucose.

The wonders of lactate

Very recently, our laboratory has made a major improvement in the use of the minimal model to assess insulin sensitivity and evaluate its importance in metabolic control. During the FSIGT, as it is practiced, there is initially a large increase in glucose concentration, due to the glucose injection, there is an endogenous finite insulin response in normal individuals, and the insulin is then increased again at 20 min after the exogenous insulin injection. We considered the question of the fate of the injected glucose before the secondary insulin injection. In the absence of a large increase in insulin, much of the glucose is disposed by an insulin-independent mechanism, a process we termed “glucose effectiveness.” Tissues which can dispose of glucose independent of a large dynamic insulin response include the brain, the liver, and red blood cells. In fact, hepatic glucose uptake is independent of the dynamic insulin response [101]. We have therefore hypothesized that the major site of glucose uptake during the FSIGT is by the liver. What is the fate of the glucose in the liver during the FSIGT? Under resting conditions, little of the glucose will be anaerobically metabolized, as there is little need for additional energy. Liver glycogen synthesis under these conditions will proceed slowly [102]. Therefore it is likely that much of the glucose will proceed down the glycolytic pathway, and exit the liver as lactate. It is possible, therefore, that the plasma lactate levels during the FSIGT reflect the ability of the liver to take up glucose in the hyperglycemic situation. The rate of liver glucose uptake is limited by the phosphorylation of glucose in the liver—which is limited by the activity of glucokinase in the liver. We have modeled the relationship between plasma
glucose and lactate during the FSIGT (Figure 15.7; [103]), and this model yields a parameter which represents the activity of glucokinase in the liver. We therefore can suggest that the activity of glucokinase can be estimated from the FSIGT. This activity is important because it is likely the major component of the glucose effectiveness.

There is now evidence that glucose effectiveness is under control by the CNS [104]; the latter concept has emerged from the use of the minimal model to assess liver glucose uptake. The latter studies, conducted in collaboration with M. Schwartz and his colleagues, utilize the minimal model in the rodent. In fact, we have collaborated with Alonso and O’Donnell and colleagues in their development of a minimal model procedure in the mouse [105]. Additionally, a genetic variant has been identified for the expression of glucokinase regulating protein, and it is possible in the future that the use of the minimal model accessing lactate will be used for identifying new risk variants for T2DM and other insulin-resistance syndromes.

**Hyperbolic relationship between insulin sensitivity and insulin secretion: the disposition index**

Diabetes, characterized by elevated blood glucose concentration, is reflective of an inability of the patient to regulate the plasma glucose concentration within normal limits. How is this regulation maintained in the normal individual?

Insulin resistance is common in Westernized societies. Yet, most insulin-resistant individuals do not have diabetes. Diabetes is prevented by the compensatory characteristics of the pancreatic \(\beta\) cells. Insulin sensitivity is reduced in many normal as well as pathophysiologic states. Among the normal states are pregnancy and puberty; pathologies include obesity and infection. Pathologies include not only T2DM, but hypertension, colon cancer, and polycystic ovarian syndrome.

Diabetes is prevented in many states of severe insulin resistance by the robust compensatory function of the cells of the pancreas. In fact, there is a predictable \(\beta\)-cell response to insulin resistance (Figure 15.8) [106]:

\[
S_I \times \text{AIR}_{\text{glucose}} = DI
\]

where the three variables are defined as in Table 15.1. This equation is a *rectangular hyperbola*, which predicts that for a given decrease in insulin sensitivity, there will be a proportionately equal and opposite increase in \(\beta\)-cell sensitivity to glucose stimulation. The disposition index defines the position of the hyperbolic curve on the sensitivity/secretion plot. If the DI is high, then the \(\beta\) cells mount a robust response to a decrease in \(S_I\); if the DI is low, such a value is reflective of unresponsive \(\beta\) cells. Subjects at risk for T2DM have a low disposition index [89,106–109]. There is evidence that the DI value is determined by genetics as well as environment [82,110,111]. The minimal model, therefore, provides a method for assessing insulin sensitivity in the intact organism by combining a relatively straightforward experimental protocol with the power of the digital computer to model real data. Software to calculate parameters has been available for some years making the methodology available to all clinical investigators. The method continues to be used in laboratories throughout the world.

**Surrogate measures of insulin sensitivity**

Laboratory methods for measuring insulin action have been shown to be accurate. However, under certain conditions application of these methods is difficult if not impractical. Some studies call for assessment of insulin action in even hundreds of thousands of subjects. Of course, the glucose clamp is not amenable to larger studies. The minimal model has been used to measure insulin sensitivity and DI in studies of populations in excess of 1000 [94,110,112], but not in multiples of that number of subjects. Are there simpler but accurate methods for use in large populations? We shall focus on two classes of
"surrogate" measures of insulin sensitivity: indices based upon fasting measures, and indices based upon the OGTT.

**Fasting insulin**

Kahn and his colleagues have reported a curvilinear relationship between fasting insulin and insulin sensitivity [113], one that is similar to stimulated insulin release. In an individual with healthy β cells, because there is a defined relationship between secretion and insulin sensitivity, it could be argued that measuring one is as good as measuring the other. Thus, if insulin sensitivity is low, fasting insulin will be elevated, and vice versa. Therefore in a group of subjects with similar β-cell function we may expect elevated insulin to reflect insulin resistance. Similarly, with healthy β cells, insulin resistance will be reflected in elevated fasting insulin, as well as elevated β-cell secretory response to stimuli such as glucose, arginine, GLP-1, or other secretagogues. It is arguably appropriate to utilize fasting insulin or a provocative index of β-cell function as a surrogate with the following caveat: fasting insulin per se, or other indices based upon the fasted state, should be considered accurate surrogates for insulin sensitivity only when comparing between or among individuals with equivalent functional activity of the pancreatic β cells [114].

The downside associated with using fasting insulin (or other related indices of insulin secretion) becomes clear if we consider comparing among groups or individuals with differing islet secretory function. Let us consider as an example comparison of fasting insulin between normal individuals and individuals with IGT. The latter group is not only insulin resistant, but is characterized by a β-cell defect of at least 50% [115–117].

Comparing the normal subjects with insulin-resistant subjects with IGT, plasma insulin concentration, or insulin response to glucose could be identical. Thus, the fasting insulin concentration will not represent an accurate reflection of insulin sensitivity comparing individuals or groups for whom cell function is not identical. In practice, individuals with similar β-cell function are not often compared; it is more usual to compare normal subjects with those at risk for disease (e.g. IGT [115–117], gestational diabetes after term [87,107], first-degree relatives of type 2 diabetic patients [118,119]).

The previous discussion demonstrates that fasting insulin per se cannot be considered an accurate surrogate for insulin sensitivity (or resistance) under most conditions. Nevertheless, there is a substantial literature using the fasting insulin value for just this purpose [120–122]. At the very least, the careful observer must consider with care reports of fasting insulin as a surrogate for insulin resistance. If there is even a latent subtle β-cell defect in a group of subjects, any reported value of fasting insulin will likely underestimate the degree of insulin resistance.

**Other surrogate measures**

Besides fasting insulin itself, there are various other surrogate measures based upon fasting values alone, which have been exploited in population studies. The first to appear in the literature was the index proposed by Dr. Peter Bennett. Others include the homeostasis model assessment index insulin resistance measure (HOMA), and the “QUICKI” [123–136]. Note that these indices are indeed similar; they represent insulin resistance as proportional to either the product or the inverse of the product of fasting insulin and fasting glucose. In fact, in nondiabetic subjects plasma glucose differs little compared to large variance in fasting plasma insulin. Therefore, in nondiabetic subjects, both HOMA index and Bennett’s index are approximately proportional to the fasting insulin concentration. As discussed, similar to fasting insulin, these indices will not reflect insulin sensitivity accurately in nondiabetic subjects with differing β-cell function.

Quon and colleagues introduced the “QUICKI” index [125]. Quon et al. equate insulin resistance to the inverse of the sum of the logarithms of fasting plasma glucose and fasting plasma insulin. The value of insulin resistance emerging from this index will depend upon the units of glucose and insulin, and the relative importance of these components will depend upon whether they are expressed in European (SI) units (mM glucose and pM insulin) or in units used in the United States (mgDL−1 glucose and μU mL−1). Also, despite the logarithmic transformation, QUICKI values exhibit nonlinear proportionality to fasting plasma insulin and will, like fasting insulin, underestimate insulin resistance in a population with a latent decrease in β-cell function. In fact, QUICKI and HOMA are formally identical to each other, as evidenced in the measures made in a large number of individuals (Figure 15.9).

The HOMA index of insulin resistance is extremely simple to calculate ((fasting glucose × fasting insulin) / 22.5). Therefore it has been used in a very large number of publications. However, it must be used with great caution. Because the fasting insulin is dependent not only upon insulin sensitivity, but also on insulin secretory function as well as metabolic clearance of insulin, the HOMA is a mixed index dependent upon all these physiologic parameters, and is not a direct measure of insulin sensitivity. The limitations of this measure have recently been documented in a series of studies [137,138]. Xiang and colleagues have shown that HOMA does not detect longitudinal changes in insulin resistance in a population at risk for diabetes [137]. It fails to detect insulin resistance associated with surgical trauma [138]. In a very recent study from our group done by Ader et al. [114], we examined whether HOMA could detect insulin resistance associated with a fat diet in the canine model. We measured insulin sensitivity with the euglycemic clamp, with the minimal model, and with HOMA. While the clamp and FSIGT values were strongly correlated, there was no significant correlation between the HOMA index and insulin sensitivity measured by either clamp or FSIGT. More significant, there was a correlation between HOMA and insulin response (Figure 15.10). Therefore based upon this study, HOMA results must be treated with a degree of skepticism, regarding insulin sensitivity per se.
Figure 15.9 Comparison of HOMA and QUICKI in normal children. Note that a mathematical conversion (log HOMA vs. 1/QUICKI) results in a 1:1 relationship between these variables. Thus, they are essentially identical measures, differing only by an algebraic manipulation.

Figure 15.10 Correlation of HOMA estimate of insulin resistance with glucose-stimulated insulin response. Source: Ader (Epub ahead of print) [114].

Thus, indices based upon fasting measurements of insulin alone or fasting insulin plus fasting glucose should not be considered accurate indices of insulin resistance comparing groups with possible inequalities in β-cell health. Because we are often comparing between individuals or groups with differing islet function, it must be concluded that fasting insulin, as well as other surrogate measures based upon fasting values alone, may not be accurate. Certainly they are inaccurate compared to measures available from laboratory procedures such as the euglycemic clamp or the minimal model.

**Insulin sensitivity based upon the OGTT**

The previous discussion suggests that “quick and dirty” indices from fasting values may not be accurate, and it is often true that laboratory procedures are impractical. Recently, it has been suggested that indices calculated from the OGTT may be useful for population studies. For decades the OGTT has been used as a screening test and a diagnostic tool for diabetes and IGT [139,140]. It might be ideal if one could classify the state of carbohydrate tolerance and simultaneously measure insulin sensitivity and β-cell function. As long ago as 1981, Voors et al. suggested the mathematical product of the 1-h glucose and the 1-h insulin concentration as "peripheral insulin resistance" [141]. Ten years later, Cederholm and Wibell expressed sensitivity as the ratio of the glucose metabolic clearance rate to the logarithmic transformed mean serum insulin concentration [141]. This latter index was later modified by Gutt et al. and shown to correlate with insulin sensitivity values derived from the euglycemic clamp [142]. More recently, Matsuda and DeFronzo applied a square root conversion of the products of fasting OGTT glucose and insulin to correct for the nonlinearity of the values [143]. Stumvoll et al. performed multiple regression of several demographic parameters (body mass index, sex, waist to hip ratio, OGTT glucose and insulin) upon sensitivity from the euglycemic hyperinsulinemic clamp test [144]. Thus, their approach included as independent variables metabolic characteristics which themselves can alter insulin action [145–148].

The OGTT-derived estimates of insulin sensitivity mentioned earlier correlated with insulin sensitivity values obtained by the euglycemic clamp technique (Table 15.2). Several of the indices were additionally tested in different ethnic groups [149–151]. Kanauchi et al. showed that Matsuda’s (r = 0.45), Gutt’s (r = 0.64), and Stumvoll’s (r = 0.53) formulas correlated modestly in a Japanese population with insulin sensitivity data from euglycemic clamps [151]. Kuo et al. reported the Matsuda index and the homeostatic model assessment equally applicable for estimation of insulin sensitivity in Chinese diabetic patients and their offspring [150]. In contrast, Chiu et al. [149] reported that the estimated indices (Matsuda and Stumvoll) correlate with clamp-derived insulin sensitivity in a wide spectrum of glucose-tolerant healthy subjects, but that they are less likely to detect differences in insulin sensitivity among different ethnic groups (Asian Americans, African Americans, Caucasian Americans, and Mexican Americans). However, the estimations of insulin sensitivity and β-cell function by OGTT-derived methods failed to reproduce the hyperbolic relation [152], which is well established for the relation between β-cell function...
Measuring insulin action in vivo

Table 15.2 OGTT-derived and fasting indices of insulin sensitivity are correlated with clamp-derived estimations of insulin sensitivity.

<table>
<thead>
<tr>
<th>Insulin sensitivity indices</th>
<th>Equation</th>
<th>n*</th>
<th>r</th>
<th>Reference</th>
</tr>
</thead>
</table>
| ISI_{0, 120} index          | \[
\frac{75000 \text{mg} - (G_k_0 - G_k_{120}) \times 0.19 \times BW}{120 \times \text{Glc}_{\text{mean}} \times \log [\text{Ins}_{\text{mean}}]}\]
|                             | 135      | 0.63 | [142] |
| Stumvoll index              | \[
0.226 - (0.0032 \times \text{BMI}) - \left(0.0000645 \times \text{Ins}_{120}\right) - (0.0037 \times \text{Glc}_{\text{mean}})\]
|                             | 104      | 0.79 | [144] |
| Matsuda index               | \[
10000 \times \sqrt{\left(\text{Ins}_0 \times \text{Glc}_0\right) \times \left(\text{Ins}_{\text{mean}} \times \text{Glc}_{\text{mean}}\right)}\]
|                             | 153      | 0.73 | [143] |
| HOMA-IR                     | \[
\frac{\left(\text{Ins} \times \text{Glc}\right)}{22.5}\]
|                             | 115      | −0.82 | [136] |
| QUICKI                      | \[
\frac{1}{(\log [\text{Ins}] + \log [\text{Glc}])}\]
|                             | 56       | 0.78 | [135] |

* n = number of subjects in respective study.

and insulin sensitivity when measured by more established methods [106].

It is clear that a correlation between proposed OGTT-derived indices and clamp-derived insulin sensitivity may exist. However, the OGTT-derived indices contain endogenous insulin response as an important term in their calculation and it cannot be excluded that reported correlations reflect islet cell response rather than insulin sensitivity per se. If it is true that OGTT-based methods reflect secretion rather than sensitivity, application of the OGTT methods to subjects with impaired \(\beta\)-cell secretory capacity (e.g., those with impaired glucose tolerance, gestational diabetes, T2DM) may underestimate insulin resistance, as postload hyperinsulinemia would be reduced and they may not be accurate for diabetic patients. Confirming this latter weakness, correlations between OGTT-based indices and clamp-based sensitivity [143,153] were lower when the indices are employed in subjects with T2DM.

Poor reproducibility of the 75-g OGTT also limits the value of the OGTT [154]. The overall test-to-test reproducibility (coefficient of variation) of the OGTT is no more than 65% [155]. The inherent variability is likely due to high day-to-day variability in gastrointestinal function (gastric emptying, absorption, incretin effect). It is well established that a delay in gastric emptying or reduced glucose absorption will result in lower glucose excursions as well as lower insulin levels [84,156,157]. In fact it has been demonstrated that gastric emptying alone accounts for 35% of the variance in peak plasma glucose after a 75-g OGTT in both healthy volunteers and patients with T2DM [84].

We have investigated the possible effects of variability in islet cell function as well as variability in intestinal glucose absorption on glucose and insulin patterns and OGTT-based indices using a computer model [158]. The glucose and insulin time series simulated for a normal healthy subject closely match published data [159]. The OGTT data simulated under these conditions indicate that ±50% changes in \(\beta\)-cell sensitivity or glucose absorption have a far greater impact on the pattern of glucose and insulin than alterations of insulin sensitivity itself. An imposed 50% reduction of insulin sensitivity caused only very minor changes in sensitivity calculated by those indices (Table 15.3). In contrast, isolated alterations of \(\beta\)-cell sensitivity and glucose absorption resulted in considerably higher or lower apparent values of insulin sensitivity from

**Table 15.3** OGTT-derived and fasting indices were calculated from the model-predicted glucose and insulin concentrations.

<table>
<thead>
<tr>
<th>Normal values</th>
<th>Imposed changes in metabolic variables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(\Delta) Insulin sensitivity</td>
</tr>
<tr>
<td></td>
<td>+50%</td>
</tr>
<tr>
<td>Relative change in insulin sensitivity values</td>
<td></td>
</tr>
<tr>
<td>ISI_{0, 120} index</td>
<td>83.2</td>
</tr>
<tr>
<td>Stumvoll index</td>
<td>0.11</td>
</tr>
<tr>
<td>Matsuda index</td>
<td>4.0</td>
</tr>
<tr>
<td>HOMA-IR</td>
<td>1.64</td>
</tr>
<tr>
<td>QUICKI</td>
<td>0.36</td>
</tr>
</tbody>
</table>

*Note:* Relative changes of the individual sensitivity values under assumption of isolated 50% changes of insulin sensitivity, insulin secretion, and glucose absorption are shown in this table.
almost all indices. While these results are from a simulation study, they suggest strongly that one must be very careful in interpreting differences in OGTT values in terms of insulin sensitivity per se because differences in metabolic parameters other than insulin sensitivity may have greater effects on OGTT. Therefore alterations in OGTT cannot be readily interpreted to reflect changes in insulin sensitivity alone, but may reflect changes in gastric emptying, insulin secretion, glucose effectiveness, and so on. Clearly, however, the OGTT (or meal tolerance tests) are far superior for assessment of insulin sensitivity than measures based upon fasting glucose and insulin alone.

**Conclusions**

The objective to accurately measure insulin sensitivity (i.e., “insulin resistance”) in patients remains an important goal. Insulin resistance is a risk factor for a plethora of chronic diseases (diabetes, cardiovascular disease, hypertension, colon cancer). Extensive efforts are under way to find genetic variants underlying insulin resistance. To study these diseases, and to intervene early in their pathogenesis, requires that we quantify insulin resistance as one important risk factor. To measure sensitivity accurately it is best to use a method that observes in some fashion the metabolic effect of insulin given intravenously. Thus, the glucose clamp and the minimal model are accurate and reproducible methods. While one must remain vigilant regarding limitations of these methods, they may be applied with confidence under a wide variety of conditions.

Unfortunately it is not possible to be equally sanguine in recommending simpler (i.e., surrogate) methods. While it may be appropriate to use fasting insulin to reflect insulin resistance in patients without any degree of β-cell failure, it is usually not possible to determine a priori whether any such defect exists. Thus, one can report fasting insulin simply as a qualitative reflection of insulin resistance. Indices that also include glucose appear to add little to the fasting insulin itself. There is strong evidence that the HOMA index does not measure insulin resistance accurately, and in fact may be a better index of β-cell response. Because the purpose of using surrogate methods is to differentiate these two physiologic functions—insulin sensitivity versus β-cell response—it is not possible to recommend the use of surrogate methods based upon fasting values, or at least be highly sensitive to the limitations of using them. Thus, if a genetic variant related to HOMA (or its logarithmic cousin, QUICKI) is identified, one cannot say with confidence the phenotypic trait related to the measured value. Caution in interpretation is the watchword.

Insulin sensitivity measures based upon oral glucose or meals are superior to fasting measure-based surrogates. One must be cognizant, however, of the possible involvement of alterations in nutrient absorption as well as gastrointestinal hormones in the values which emerge. Hormones such as GLP-1 alter insulin action [160], and such effects will alter the sensitivity measures emanating from the oral tests.

Clearly more work is justified to identify new and novel methods to obtain accurate and precise measures of insulin action in vivo from simple-to-perform tests. Among others, we continue this quest in our own Institute.

**Acknowledgments**

RNB was supported by the National Institutes of Health (DK27619 and DK29867). We are grateful to Anne E. Sumner, MD, for the data in Figure 15.6.

**References**


65 Hamilton-Wessler M, Buchanan TA, Sogaard B, et al.: Lesser equimolar effect in humans of fatty acid acylated insulin is not due to increased liver extraction. Diabetes 2000;49:A110.
86 Amoah AG, Owusu SK, Schuster DP, Osei K: Pathogenic mechanism of type 2 diabetes in Ghanaians: the importance of beta


116 Weyer C, Bogardus C, Pratley RE: Metabolic characteristics of individuals with impaired fasting glucose and/or impaired glucose tolerance. Diabetes 1999;48:2197–2203.


126 Duncan GE, Hutson AD, Stacpoole PW: QUICKI is not a useful and accurate index of insulin sensitivity following exercise training. *Journal of Clinical Endocrinology & Metabolism* 2002;87:950–951.


149 Chiu KC, Martinez DS, Yoon C, Chuang LM: Relative contribution of insulin sensitivity and beta-cell function to plasma glucose
and insulin concentrations during the oral glucose tolerance test. *Metabolism* 2002;51:115–120.


154 Sievenpiper JL, Jenkins DJ, Josse RG, Vuksan V: Dilution of the 75-g oral glucose tolerance test improves overall tolerability but not reproducibility in subjects with different body compositions. *Diabetes Research and Clinical Practice* 2001;51:87–95.


CHAPTER 16

Protein metabolism in health and diabetes

Elena Volpi1 and Jared M. Dickinson2

1 Institute for Translational Sciences, University of Texas Medical Branch, Galveston, TX, USA
2 School of Nutrition and Health Promotion, Arizona State University, Phoenix, AZ, USA

Key points

• Insulin is the most powerful protein anabolic hormone over the short term. In physiologic conditions, the anabolic effect of insulin on protein metabolism occurs during feeding, when insulin secretion promotes dietary amino acid storage in sensitive tissues, mainly skeletal muscle.

• The mechanisms by which insulin stimulates protein anabolism are complex and involve both an increase in protein synthesis and a decrease in proteolysis. These mechanisms are triggered by insulin binding to its receptor and are mediated by Akt and mTORC1 signaling.

• Insulin stimulation of protein synthesis is dependent upon amino acid availability and the hormone’s ability to enhance endothelial-dependent vasodilation. In fact, endothelial dysfunction and/or a reduction in nutrient (amino acid) availability induce resistance to the stimulatory effect of insulin on protein synthesis (anabolic resistance).

• Insulin exerts differential protein anabolic effects on its target tissues. In skeletal muscle, insulin enhances protein synthesis unless amino acid availability decreases or blood flow does not increase, in which case a reduction in proteolysis prevails. In the splanchnic region, insulin influences the synthesis of some export liver proteins. Insulin also reduces skin protein breakdown and overall skin protein turnover with no effect on the net protein balance in healthy subjects, while it stimulates wound protein anabolism by reducing protein breakdown. At the whole body level insulin reduces protein breakdown despite its stimulatory effect on muscle and liver proteins. This apparent discrepancy may be explained by methodological limitations of the model for measurement of whole body protein turnover and/or by the possibility that insulin may also decrease the synthesis of other unmeasured proteins, offsetting the stimulatory effect on muscle protein synthesis.

• The effect of insulin on protein metabolism changes according to the pathophysiologic setting. In some conditions the stimulatory effect on protein synthesis is reduced (e.g. aging), while in others the antiproteolytic effect is enhanced (e.g. exercise) or reduced (e.g. type 2 diabetes).

Introduction

Over the last two decades it has become clear that insulin, the well-known regulator of glucose and fat metabolism, plays also a critical role in the regulation of protein metabolism. Specifically, insulin is the most powerful protein anabolic hormone over the short term. The anabolic effect of insulin on protein metabolism occurs during feeding, when the large insulin secretory bursts promote storage of dietary amino acids. It is important to underscore that, unlike glycogen for glucose and triglycerides for fatty acids, there is no inactive storage moiety for amino acids, which must be incorporated into functional proteins. Skeletal muscle contractile proteins are the largest functioning storage system for essential amino acids [1]; consequently they are a major target of the protein anabolic effect of insulin. An adequate basal insulin concentration as well as normal insulin signaling is also required to prevent rapid protein loss. Conditions of insulin deficiency (untreated type 1 diabetes), or severe insulin resistance (e.g. critical illness) are characterized by the release of large amounts of amino acids from skeletal muscle, which leads to rapid muscle wasting. Recent observational studies have also highlighted that type 2 diabetes accelerates sarcopenia, the loss of skeletal muscle mass and function with aging. These clinical observations highlight the importance of insulin for skeletal muscle homeostasis and function. However, other proteins and tissues are also significantly impacted by the protein anabolic effect of insulin, particularly in the liver and skin.

The mechanisms by which insulin regulates protein metabolism are not completely understood. While insulin action has been extensively examined in vitro, in cell systems, and in animal models, it is very difficult to determine if those mechanisms are also involved in the physiologic regulation of protein metabolism in vivo, in humans. The main reason is the complexity of the system. Insulin can potentially affect a wide variety of proteins in a variable manner. The situation is further complicated by the physiologic setting in which insulin...
is normally secreted: the prandial state. During feeding, the increase in insulin concentration occurs concomitantly with major changes in substrate availability. It is thus very difficult to isolate the direct effects of insulin on protein turnover from those due to changes in amino acid and energy availability. Moreover, changes in insulin concentration also induce changes in tissue blood flow, which can significantly affect the hormone’s action on protein synthesis and breakdown. This chapter will provide an overview of the known mechanisms of insulin action on protein metabolism in humans utilizing an integrative approach.

**Integrative physiology of insulin action on protein metabolism**

Insulin impacts a variety of physiologic functions that are regulated through the activity of an intricate array of cellular mechanisms. These mechanisms are extremely important for insulin to regulate protein turnover in metabolically active tissues, particularly skeletal muscle and the liver. In addition, insulin also triggers a series of equally important cellular events in the endothelium that allow it to exit the circulation and exert its actions on the target tissues.

This section focuses on the mechanisms that regulate the physiologic responses of protein metabolism to insulin in humans. We will follow the fate of insulin starting with its exit from the circulation to reach the target tissues, and discussing the role of insulin-induced vasodilation on protein anabolism. We will then examine the direct insulin effects on target tissues and conclude with the overall effects of insulin at the whole body level. Figure 16.1 provides a simplified depiction of the effects of insulin on the endothelium and the muscle tissue, which is the major target of insulin action on protein metabolism.

**Insulin transport from the circulation to the target tissues**

To exert its physiologic actions on the target tissues, insulin must first leave the circulation by crossing the endothelial barrier. This process involves changes in local blood flow (i.e., microvascular blood flow) and the ability for insulin to exit the circulation, enter the interstitial space, and bind to the target tissues’ receptors. This series of events is critical for skeletal muscle to mount a protein anabolic response to hyperinsulinemia.

Extensive work from Barrett and colleagues has highlighted that a major rate-limiting step for insulin to exert its physiologic actions on skeletal muscle is its transport from the circulation to the interstitial space [2]. The model proposed suggests the existence of a saturable transendothelial transport mechanism [2,3]. Two insulin-dependent processes work in concert to regulate insulin transport: (1) the binding of insulin to its receptor and the subsequent internalization of this complex into the endothelial cell [3], and (2) the stimulation of insulin signaling in the endothelial cell [2]. More specifically, insulin-dependent activation of the phosphoinositide 3-kinase (PI3K)-Akt and mitogen activated protein kinase (MAPK) pathways appear to have an important role in promoting insulin uptake and vesicular intracellular trafficking of insulin bound to its receptor [3]. In other words, insulin appears to stimulate its own transport across the endothelium in a positive feedback manner that requires an intact insulin signaling cascade [2]. In addition, several other proteins, including Caveolin-1 [4], appear to have an important role in supporting insulin uptake into the endothelium and the subsequent transendothelial vesicle trafficking of insulin through the endothelial cell into the interstitial space. These complex mechanisms involved in the regulation of insulin transport from the circulation to the extracellular space are a major rate-limiting step for insulin to exert its physiologic actions on the target tissues, particularly skeletal muscle [3].

**Insulin-induced vasodilation**

Insulin transport through the endothelium triggers a powerful vasodilatory response, which increases microvascular nutritive flow to the target tissues. The mechanism involves activation of the endothelial nitric oxide synthase (eNOS) via protein kinase B (PKB or Akt) dependent phosphorylation [5]. In turn, activation of eNOS stimulates the release of nitric oxide (NO) in the endothelial cells causing dilation of the peripheral vasculature [6].

A major role for insulin-induced vasodilation in the protein anabolic response to the hormone was initially hypothesized to account for otherwise unexplainable differences in the results of earlier human studies examining the metabolic mechanisms by which insulin stimulates skeletal muscle protein growth [7]. Recent mechanistic studies in humans have demonstrated a causal relationship between the vasodilatory action of insulin and its anabolic effect on skeletal muscle protein anabolism. These experiments involved the local insulin infusion technique with euglycemic clamp to isolate the direct effect of insulin from that of amino acid and glucose availability, as well as NO synthase inhibition with NG-monomethyl-L-arginine (L-NMMA) [8]. When insulin was infused alone it increased skeletal muscle protein synthesis and net balance. However, co-infusion of L-NMMA with insulin blocked not only the insulin-induced endothelial dependent vasodilation but also the insulin-stimulated increase in skeletal muscle protein anabolism [8]. Conversely, in a condition of protein anabolic resistance to insulin that is accompanied by endothelial dysfunction, sarcopenia of aging [9], the stimulatory effect of insulin on muscle protein anabolism can be restored by normalizing vasodilation via co-infusion of insulin with sodium nitroprusside, a NO donor [9,10]. These data demonstrate that increases in blood flow and microvascular perfusion are critical for insulin to exert its anabolic effects on skeletal muscle protein metabolism. As we will discuss more in detail later, insulin’s inability to trigger these endothelial events and increase microvascular flow to skeletal muscle, as observed in
Molecular mechanisms of insulin action on protein metabolism

After its transport out of circulation, insulin binds to the receptor in the target tissues. The metabolic actions of insulin on the target tissues are produced through insulin-dependent activation of key intracellular signaling pathways. The pathway involved in the stimulation of muscle protein anabolism is intricately connected with the signaling pathway involved in glucose uptake. Thus, we will describe later the cellular mechanisms by which insulin regulates these two very well-defined physiologic processes. We will focus on the effects of insulin on skeletal muscle, which is the tissue most studied in humans.

Insulin increases glucose uptake in skeletal muscle through the glucose transporter type 4 (GLUT4). This process requires the translocation of GLUT4 from its intracellular storage site to the cell surface. The initial event signaling GLUT4 translocation is the binding of insulin to its cell surface tyrosine kinase receptor.
This binding triggers the phosphorylation of insulin receptor substrate 1 (IRS-1) and the subsequent phosphorylation and activation of PI3K. Activation of PI3K eventually triggers the phosphorylation of Akt (also known as protein kinase B [PKB]), which is phosphorylated by both phosphoinositide-dependent kinase 1 (PDK1) and the mammalian target of rapamycin complex 2 (mTORC2) [11]. The phosphorylation of Akt represents a critical step in insulin signaling, and is a branching point for signaling related to insulin-stimulated protein anabolism.

One of the primary downstream targets for Akt in the cellular regulation of insulin-stimulated glucose uptake appears to be the Akt substrate of 160 kDa (AS160) [12]. AS160 seems to play a role in retaining GLUT4 to its intracellular vesicles through its interaction with specific GTPases [11]. Phosphorylation of AS160 by Akt is thought to release the inhibitory effects of AS160 on GLUT4 translocation, thus allowing GLUT4 to translocate to the cell surface with support of other regulatory proteins [11]. Upon removal of the insulin signal, a majority of the GLUT4 will relocate to the intracellular vesicles. In fact, in the absence of a significant insulin signal (i.e., in the fasting state) it is estimated that greater than 95% of the GLUT4 may be localized intracellularly, while only a small fraction is found on the cell surface [13]. The inability of insulin to trigger the translocation of GLUT4 to the surface of skeletal muscle cells (and adipose tissue) represents a key component in the development of insulin resistance.

The other major pathway that branches off downstream of Akt is the mTORC1 signaling pathway, which is involved in the regulation of muscle protein synthesis and anabolism. Originally, it was thought that Akt could signal to mTORC1 by phosphorylating and inhibiting TSC2, which is a direct upstream negative regulator of mTORC1. However, more recent data suggest that Akt increases mTORC1 activity by phosphorylating proline-rich-Akt/PKB substrate 40Kda (PRAS40) [14], which binds to mTOR and is a negative regulator of mTORC1 activity. The Akt-dependent phosphorylation of PRAS40 facilitates the release of PRAS40 from mTOR (perhaps via 14-3-3 binding to PRAS40) [14], resulting in increased mTORC1 activity.

Protein anabolism is largely regulated by the mTORC1 signaling pathway in humans [15,16]. It is very well understood that an activation of mTORC1 signaling triggers an increase in protein synthesis and also attenuates muscle protein breakdown, although the latter mechanism is less well characterized in humans. Collectively, the increase in protein synthesis and the decrease in protein breakdown following the activation of mTORC1 signaling creates a cellular environment that favors a positive net protein balance. In humans, these effects of insulin have been described in skeletal muscle [8–10]. The mTORC1 signaling pathway is an anabolic hub for a variety of other anabolic stimuli which interact with insulin to enhance muscle growth. These include nutrients, primarily amino acids, and contraction. The remainder of this section will focus on the cellular mechanisms through which insulin stimulates protein metabolism and activates mTORC1 signaling.

The two best-characterized downstream targets of mTORC1 are the eukaryotic initiation factor 4E binding protein 1 (4E-BP1) and p70 ribosomal S6 kinase 1 (S6K1). Phosphorylation of 4E-BP1 by mTORC1 induces the release of eIF4E from 4E-BP1 and the subsequent formation of the eIF4F translation initiation complex [17]. Phosphorylation of S6K1 by mTORC1 further enhances translation initiation, primarily that of terminal oligopyrimidine (TOP) mRNAs [18]. Activation of S6K1 also contributes to increased translation elongation via the phosphorylation of the eukaryotic elongation factor 2 (eEF2) kinase and the subsequent activation of eEF2 [17]. In addition to the role of S6K1 in mediating translation initiation and elongation, phosphorylation of S6K1 negatively regulates insulin signaling. In particular, activation of S6K1 triggers the phosphorylation of IRS-1 at specific amino acid residues, which inhibits IRS-1 and promotes insulin resistance [19]. The existence of this negative feedback loop suggests that chronic activation of mTORC1 signaling, perhaps through excess food intake, may be a mechanism that contributes to the development of insulin resistance [20].

Activation of the insulin signaling cascade also reduces protein breakdown by stabilizing lysosomes and reducing the activity of the ubiquitin-proteasome pathway [21]. In muscle, the effects of insulin on protein breakdown are in part mediated through the activation of Akt and subsequent phosphorylation of the Forkhead Box O (FOXO) 3 transcription factor. Specifically, FOXO3 is responsible for the transcription of two well-known E3 ubiquitin ligases, MuRF1 and atrogin-1, and the phosphorylation of FOXO3 by Akt inhibits the movement of FOXO3 into the nucleus [22]. Consequently, FOXO3 cannot stimulate the transcription of MuRF1 and Atrogin-1 and the ubiquitin proteasome system is attenuated. In addition, recent data indicate that FOXO3 may also mediate the transcription of genes involved in autophagy, including LC3 and Bnip [23]. Thus, phosphorylation of FOXO3 by insulin through the activation of Akt may also attenuate muscle protein breakdown via the inhibition of autophagy. Furthermore, activation of mTORC1 by insulin has also been linked to the inhibition of the autophagy system. In particular, mTORC1 phosphorylates and inactivates the ULK1 complex, which inhibits the initiation step of autophagosome formation and subsequent breakdown of proteins by autophagy [24]. Collectively, these findings highlight that there is an intricate network of cellular regulatory pathways that, when stimulated by insulin, can increase initiation of translation and protein synthesis, and also reduce protein breakdown through the inhibition of the two major proteolytic systems.

**Effect of insulin on tissue amino acid availability**

Control of tissue amino acid availability is an additional mechanism that appears to be necessary for activation of protein anabolism by insulin. This is particularly important in skeletal muscle. Under physiologic conditions, insulin exerts two main effects on amino acid availability.
First, insulin enhances amino acid delivery to the muscle via stimulation of vasodilatation (see earlier section “Insulin-induced vasodilatation”). An increase in blood insulin concentration increases the amount of tissue that is exposed to amino acids via endothelium-dependent vasodilation. In turn, the stimulation of muscle perfusion increases the flux of amino acids delivered to the tissue (i.e., the product of blood amino acid concentration by blood flow to the tissue) [8,25,26]. The increase in amino acid delivery can occur even in conditions of low blood amino acid concentrations, such as during insulin infusion in the fasting state, as long as the blood amino acid concentrations do not decrease below the lower normal value [7]. However, it is also important to underscore that vasodilation alone, in the absence of hyperinsulinemia, is unable to stimulate mTORC1 signaling and muscle protein synthesis [27]. In other words, vasodilatation and increased amino acid delivery are necessary but not sufficient mechanisms for insulin to stimulate protein synthesis and anabolism. This indicates that insulin signaling must also be activated in the target tissue to stimulate protein anabolism.

Second, insulin stimulates amino acid uptake and transport into a variety of tissues including skeletal muscle [7], and is directly involved in the regulation of amino acid transporter mRNA and protein expression [28]. This process is emerging as a major regulator of skeletal muscle protein homeostasis [29]. Specifically, activation of PI3K by insulin stimulates downstream activation of Akt, which, in turn, activates mTORC1 [14]. Once activated, mTORC1 not only stimulates protein synthesis via S6K1 [18], but also increases the expression of amino acid transporters via ATF4 in cell systems [28]. Thus, an impaired Akt/mTORC1 signaling can decrease amino acid transport into the muscle cell and blunt muscle protein synthesis. Kinetic studies using stable isotope methodologies have highlighted that hyperinsulinemia does indeed increase amino acid uptake from blood into skeletal muscle in healthy young humans, that this process appears to be a limiting factor for the direct measurement of the net anabolic effect of insulin on muscle protein synthesis [7].

**Effects of insulin on protein metabolism in specific tissues**

The effects of insulin on protein metabolism vary in the different tissues and organs. In some tissues, insulin exerts its anabolic action primarily via increased protein synthesis, while in others its antiproteolytic effect prevails. We will discuss below the effects that insulin exerts on the tissues that contribute the most to protein metabolism at the whole body level: skeletal muscle, splanchnic tissues, and skin.

**Skeletal muscle**

Skeletal muscle contains approximately 50% of the body protein [1]. Insulin is an important anabolic hormone in humans because it physiologically increases net accretion of skeletal muscle proteins in healthy, young adults [7,30,31]. The net anabolic effect of insulin on skeletal muscle proteins has been unequivocally reported in the literature [7,30–41]. However, the mechanisms by which insulin enhances muscle protein anabolism in humans have been the matter of intense discussion for more than two decades. Some of the studies reported a stimulatory effect of insulin on muscle protein synthesis [7,31,32,37,41,42], while others reported primarily a decrease in muscle protein breakdown [7,30,33–36,40,41]. There are several possible explanations for these discrepancies between human studies.

First, methodological issues should be considered. Muscle protein turnover is typically measured on the leg or the forearm with stable isotope tracer methodologies using the arteriovenous balance technique and/or the precursor-product method [43]. Selection of the sampling site, leg or forearm, is critical for the arteriovenous balance technique. This method implies the measurement of the balance of unlabeled amino acids and amino acid tracers across the selected limb, assuming that amino acid metabolism occurs mostly in the limb skeletal muscle. However, it is possible that the relative contribution of non-muscle tissues (skin, bone, bone marrow) to leg or forearm amino acid turnover is not the same. Insulin decreases skin protein turnover [44] (see later section on “Skin”). Bone marrow proteins are likely to turnover faster than muscle proteins because of the high cellular turnover, but no data are available on the effects of insulin on this tissue. Thus, the sampling site may have been partly responsible for some of the differences reported between studies. Selection of the arteriovenous balance model to calculate muscle protein turnover is even more important and can significantly affect the conclusions of the study. There are three fundamental models available to measure human muscle protein synthesis [43]: the two arteriovenous balance models, two-pool and three-pool, and the precursor-product model. The two-pool model estimates muscle protein synthesis and breakdown using the amino acid enrichments and concentrations in the artery and vein of the selected limb. These parameters are based on the extraction of the labeled amino acid from the artery, the appearance of unlabeled amino acid from the muscle into the vein, and the net arteriovenous amino acid difference [43]. Thus, this model provides data regarding the kinetics of blood amino acids across the leg or forearm, but cannot measure the intracellular recycling of those amino acids that are released from protein breakdown and re-incorporated into proteins without appearing in the blood. In other words, the two-pool arteriovenous method allows for the measurement of the effect of treatments on the net kinetics of blood amino acids across the selected limb, while not offering any insight into the actual intracellular amino acid kinetics. The three-pool model is an expansion of the two-pool model and relies not only on the measurement of amino acid enrichments and concentrations in the arterial and venous blood, but also on the direct measurement of the amino acid enrichments in the free tissue water. This allows for the direct measurement of the intracellular amino acid
kinetics, including recycling from breakdown to synthesis. In addition, it also allows for calculation of the amino acid transport rates from artery into the muscle tissue, and from the muscle tissue into the venous blood. Many of the human studies on the effects of insulin on muscle protein turnover have utilized the two-pool method [7,8,30,33–40,45,46], but most recently several studies have utilized the three-pool method either alone [31] or in conjunction with the two-pool and/or the precursor product method [7,8,45,46]. In those studies in which more than one model was used the results of the two- and three-pool model tended to be consistent, while the results of the arteriovenous balance methods and the precursor-product method not always agreed possibly due to the different timing of the measurements [7]. Samples for the arteriovenous methods are in fact collected in a shorter time frame as compared to samples for the precursor-product method. The latter involves measurement of the incorporation rate of an amino acid tracer into muscle proteins and provides an integrated view of the overall effect of insulin on muscle protein synthesis over several hours.

A second and, possibly, more important issue to be considered is differences in study design that may have differentially emphasized the complex mechanisms by which insulin exerts its action on muscle protein metabolism. Specifically, a review of the human studies on the effects of insulin on skeletal muscle protein turnover suggests that such conflicting results could be entirely explained by differences in amino acid availability during insulin infusion. The studies reporting an increase in skeletal muscle protein synthesis as the primary mechanism of the insulin-induced muscle protein anabolism also reported an increase in amino acid availability [7,31,32,37–42]. Conversely, the studies that found a significant reduction in skeletal muscle protein breakdown with no significant changes in protein synthesis during hyperinsulinemia also reported a decrease or no change in amino acid availability [7,30,33–36,41]. The differences in amino acid delivery and availability were due to differences in amino acid concentrations and blood flow. The changes in amino acid concentrations were determined by the modality of insulin infusion, and/or by the concomitant infusion of exogenous amino acids, or lack thereof. Specifically, systemic insulin infusion dramatically decreases blood amino acid concentrations [30,40] unless amino acids are replaced by exogenous infusion [30,34,37,39,40,47,48]. Conversely, local insulin infusion in a leg or a forearm allows for the exposure of muscle tissue to relatively high insulin levels while avoiding a major reduction in blood amino acid concentration [31,33]. As discussed in detail earlier, a fundamental variable that significantly affects the response of skeletal muscle protein synthesis to insulin is blood flow. Under physiologic conditions of normal amino acid concentrations, insulin has been reported to enhance microvascular flow [26,49], amino acid delivery to the muscle, and amino acid transport into the myofibers [7,8,31], with a resulting stimulation of skeletal muscle protein synthesis [7,8,31].

Third, it is also important to highlight that skeletal muscle protein metabolism is less sensitive to insulin as compared to glucose or fat metabolism, so that the stimulatory effect of insulin on skeletal muscle protein synthesis occurs only at or beyond physiologic postprandial levels of insulin concentrations (approximately 50 μU mL⁻¹ or 300 pmol L⁻¹) [7,31,37]. Lower insulin doses, such as doubling the basal insulin concentration (approximately 20–25 μU mL⁻¹ or 120–150 pmol L⁻¹), have no effect on protein synthesis [7,46] and may [46] or may not [7] decrease muscle protein breakdown.

Fourth, it is important to recognize that measures of skeletal muscle protein metabolism represent the combined result of the turnover of individual proteins. Animal studies have reported that insulin can differentially affect the synthesis of individual proteins in the skeletal muscle [50]. Future studies will have to explore the potential differential effects of insulin on individual skeletal muscle protein turnover.

**Splanchnic tissues**

Insulin may affect the protein turnover rates of splanchnic tissues, particularly the liver. However, most of the data rely on the measurement of export liver proteins because it is very difficult to measure parenchymal liver protein turnover in humans. The liver arteriovenous amino acid balance methods are not applicable in healthy humans due to the difficulty of obtaining portal venous blood. One study performed using the splanchnic arteriovenous balance technique with amino acid tracers in humans, that is, a technique that involves measurement of the amino acid and amino acid tracer balance across the entire splanchnic region (including both liver and intestine), did not report a major effect of insulin on these tissues [40]. However, stable isotope tracer methods cannot be accurately utilized with the splanchnic arteriovenous balance method because the liver can oxidize all circulating amino acids, thereby impeding the correct estimation of parenchymal liver protein synthesis. The arteriovenous model assumes that the disappearance of an amino acid occurs only via synthesis, which is not the case when amino acids are oxidized by the liver. The precursor-product method is also not applicable due to the invasiveness of the liver biopsy procedure. However, the precursor-product method can be used to measure the effect of insulin on the synthesis of liver export proteins, such as albumin and acute phase proteins. Using this technique it has been shown that insulin increases albumin synthesis in normal [51] and insulin-deficient diabetic subjects [52] while decreasing fibrinogen [51,52] and antithrombin III [51] synthesis. Increased amino acid availability enhances the response of albumin synthesis to insulin in normal humans [53], suggesting that anabolic mechanisms similar to those discussed for skeletal muscle are also present in the liver. Since insulin also stimulates the transcapillary escape of albumin, it has been hypothesized that the insulin-induced increase in albumin synthesis is another important anabolic mechanism to spare dietary amino acids from oxidation and deliver them to the peripheral tissues in form of albumin [53].
The negative effect of insulin on fibrinogen and antithrombin III synthesis may also explain why lack of insulin or insulin resistance accelerates fibrinogen synthesis and atherosclerosis in diabetes.

Skin

The skin makes up approximately 20% of the entire body mass and contributes significantly to whole body protein metabolism because of its relatively rapid protein turnover rate. However, data on the effects of insulin on skin physiology are sparse and obtained mainly in animals. Using stable isotope tracers and a rabbit model to measure the arteriovenous amino acid balance of the ear, which mainly represents skin protein metabolism [54], it has been shown that the skin net amino acid balance is not significantly different from zero. Insulin decreases the overall skin protein turnover [55]. Specifically, insulin decreases skin protein synthesis and breakdown to the same extent, resulting in no significant changes in net protein balance. Since skin protein turnover has been shown to be dependent upon recycling of amino acids released by skin protein breakdown, it is likely that the main effect of insulin on skin protein turnover is initially a reduction in skin protein breakdown, which is then followed by a secondary decrease in skin protein synthesis. Interestingly, insulin becomes anabolic for skin proteins during wound healing [55]. Also this effect is attributable to a reduction in skin proteolysis, rather than an increase in skin protein synthesis.

Effects of insulin on whole body protein turnover

The effect of insulin on protein turnover at the tissue level is clearly anabolic. However, the basic mechanisms vary from tissue to tissue, with some tissues responding with an increase in protein synthesis and others with a decrease in protein breakdown. When the effect of insulin on protein metabolism is measured at the whole body level, the predominant anabolic effect is a reduction in protein breakdown rather than an increase in protein synthesis. This may seem surprising, given the results of the studies in human skeletal muscle and those in animals and cell systems indicating that insulin stimulates protein synthesis (see earlier sections on integrative physiology of insulin action and effects of insulin on protein metabolism). However, several studies performed measuring whole body leucine turnover in humans during systemic insulin infusion with euglycemic clamp have reported that insulin reduces whole body protein breakdown [56–58]. This effect is dose-dependent. When insulin concentrations are within the physiologic postprandial range protein breakdown as well as leucine oxidation are significantly reduced [56–60]. The insulin-induced decrease in protein breakdown reduces amino acid availability for synthesis. Conflicting results have been reported when amino acids are administered during hyperinsulinemia, with some studies indicating an increase in whole body protein synthesis [53,61] while others report no changes [59,60].

The results of the whole body protein metabolism studies are at odds with those obtained in skeletal muscle and liver showing that insulin stimulates protein synthesis (see previous section “Effects of insulin on protein metabolism in specific tissues”), particularly if we consider that muscle and liver proteins are quantitatively important contributors to whole-body protein turnover. Limitations of the whole body technique to measure protein synthesis may provide a likely explanation for this inconsistency. Specifically, when leucine is utilized as a tracer, whole body protein synthesis is indirectly calculated as the difference between protein breakdown and leucine oxidation. However, the measurement of breakdown and oxidation is not precise due to several issues [43]. The most important is a main assumption underlying the tracer dilution model that each tissue contributes to the dilution of the leucine tracer in the blood proportionally to the individual tissue protein breakdown rate. However, this assumption is not valid because insulin exerts differential effects on protein breakdown in the various tissues (see previous section on effects of insulin on protein metabolism). In addition, insulin can affect amino acid transport [9], resulting in a differential release of free amino acids by the various tissues, which further impacts blood leucine enrichment. Calculation of leucine oxidation also requires the assumption that blood leucine enrichment or blood alpha-ketoisocaproate enrichment represents the enrichment of the true precursor for oxidation, which, for the reasons just mentioned, is probably not correct. Thus, the estimation of whole body protein synthesis from the difference between two imprecise measures, whole body protein breakdown and amino acid oxidation, can amplify the error and lead to incorrect conclusions. However, it is also possible that insulin inhibits the synthesis of enough, unmeasured proteins to offset its stimulatory effect on skeletal muscle and liver protein synthesis. Future studies focusing on the effect of insulin on the synthesis of individual proteins are warranted.

Pathophysiology of insulin action on protein metabolism

In the following sections we provide a brief summary of the changes that common conditions induce on the response of protein metabolism to insulin.

Aging

There is growing evidence [9,10,42,45] that healthy aging per se reduces the muscle protein anabolic action of insulin in nondiabetic older persons. This anabolic resistance to insulin likely contributes to sarcopenia, the loss of skeletal muscle mass and function with aging, and frailty in older adults [62]. The skeletal muscle anabolic resistance to insulin is associated with decreased insulin-induced vasodilation and a reduction
in amino acid delivery and transport into the muscle, which prevent the physiologic activation of mTORC1 signaling and protein synthesis [9,10,42,45]. This is a true insulin resistance, because it is overcome by supraphysiologic insulin concentrations [45]. Endothelial dysfunction is its main driver, as pharmacologic vasodilation with sodium nitroprusside restores the anabolic response of muscle protein synthesis to insulin in healthy older subjects [10]. However, vasodilation alone, in the absence of hyperinsulinemia, does not stimulate mTORC1 and muscle protein synthesis [27]. This suggests that skeletal muscle insulin signaling must also be activated to stimulate protein anabolism.

**Physical activity**

Physical activity can influence the effect of insulin on protein metabolism. Aerobic exercise can overcome the insulin resistance of skeletal muscle protein metabolism in older adults by improving endothelial function and mTORC1 signaling [63,64]. Vice versa, insulin enhances the positive effect of resistance exercise on muscle protein anabolism by reducing the post-exercise increase in muscle protein breakdown [65].

**Protein metabolism in diabetes**

Protein metabolism is significantly altered by diabetes. We provide in the following sections a summary of the changes that occur in type 1 and type 2 diabetes as well as in critical illness, a condition of severe insulin resistance.

**Type 1 diabetes mellitus**

Uncontrolled type 1 diabetes has profound catabolic effects on nitrogen balance and protein metabolism. Immediately after the discovery of insulin in the 1930s, Atchley et al. demonstrated how insulin deficiency severely impacts nitrogen metabolism by inducing nitrogen loss and resulting in a profoundly negative nitrogen balance, an indication of net protein catabolism [66]. More recently, studies on leucine turnover have shown that the net protein catabolism induced by insulin deprivation is due to an increase in whole body protein breakdown [67–73]. Type 1 diabetes was also found to be associated with increased leucine oxidation during insulin deficiency [67–72]. A higher oxidation rate is indicative of an increased net loss of essential amino acids [67–72]. Surprisingly, in the same insulin-deprived type 1 diabetic patients the nonoxidative utilization of leucine (nonoxidative leucine disposal), which is an indicator of whole body protein synthesis, was reported to be normal or increased [67–72]. The puzzling results of a normal or quasi-normal whole body protein synthesis during insulin deficiency in the context of severe nitrogen loss may be due to a methodological artifact, because the protein synthesis values were indirectly calculated as the difference between whole body leucine flux and leucine oxidation. However, skeletal muscle protein synthesis, directly measured with tracers, was reported to be normal in insulin-deficient type 1 diabetes patients [69] or decreased in streptozotocin diabetic rats [74]. Thus, if the observed increase in whole body protein synthesis with insulin deficiency is real and not due to a methodological artifact, it raises the question of where this increase occurs. Several studies have highlighted that insulin can differentially affect the synthesis rate of liver secretory proteins [51,52,75–77] while reducing splanchnic protein synthesis [78]. Specifically, synthesis of the acute phase response proteins has been reported to increase during insulin deficiency [52,75–77].

**Type 2 diabetes mellitus**

The main hallmark of type 2 diabetes mellitus is a global insulin resistance which includes a reduction in skeletal muscle insulin signaling [79]. Thus, type 2 diabetes has the potential to impact protein homeostasis. Specifically, type 2 diabetes is characterized by impaired insulin sensitivity. Particularly relevant for protein metabolism is the reduction in insulin signaling through the PI3K pathway that occurs with type 2 diabetes. The PI3K pathway is essential to transport insulin across the endothelium to the target tissues. Moreover, PI3K is an upstream stimulator of Akt, which activates mTORC1. In turn, mTORC1 stimulates amino acid transporter expression via ATF4 (see earlier section on integrative physiology of insulin action). Thus, impaired insulin signaling could decrease amino acid transport into the muscle cell and blunt muscle protein synthesis. A very limited number of studies have addressed the effects of type 2 diabetes on protein metabolism. In young and middle-aged adults, poorly controlled type 2 diabetes has been reported to increase whole body [80] and muscle protein turnover [81], and induce a net nitrogen loss [82]. However, other studies have not confirmed those findings [83–85]. Poorly controlled type 2 diabetes in young adults does not significantly impact the basal muscle protein synthesis rate, while it increases leucine nitrogen flux and protein breakdown [81,85]. These changes can be corrected by hyperinsulinemia [81,86]. Data in older adults with type 2 diabetes are even scantier. One study reported no differences between healthy and type 2 diabetes older adults in the response of muscle protein synthesis to a mixed meal [87]. However, the meal contained a very large amount of protein (more than 100 grams), and the results were also confounded by the use of various glucose-lowering medications in some of the subjects. Independent observational studies have recently reported that type 2 diabetes accelerates the loss of muscle mass and strength in older persons, even after adjustment for comorbidities and other factors [88–91]. Another cohort study reported that insulin sensitizers may slow down the rate of decline of muscle mass in older men with type 2 diabetes [92], suggesting an important role for insulin resistance in this process. Altogether these data suggest that type 2 diabetes may have negative effects on protein metabolism particularly in older adults.
Critical illness

Critical illness is associated with severe insulin resistance resulting in hyperglycemia. However, this insulin resistance does not translate into resistance to the protein anabolic action of insulin, as insulin infusion can still suppress whole body protein breakdown [93] and increase muscle protein synthesis [94] in critically ill patients.

Conclusions

Insulin is a major protein anabolic hormone that promotes dietary amino acid storage in the fed state and exerts an anticasabolic role in the basal state and during critical illness. The mechanisms by which insulin exerts its anabolic action are complex and challenging to study because insulin is normally secreted during feeding. Thus, it very difficult to isolate the insulin effect from that of increased nutrient and energy availability. Moreover, methodological issues further complicate the precise measurement of the effects of insulin on protein turnover.

In general, the effects of insulin are tissue specific, increasing protein synthesis in some tissues, such as skeletal muscle, and reducing protein breakdown in others, such as the skin. In several instances insulin’s effect is also protein specific. For example it increases the synthesis of some liver export proteins, such as albumin, and may reduce the synthesis of others, such as fibrinogen. These effects can be magnified or reduced by specific pathophysiologic conditions, such as exercise, diabetes, or critical illness.

Acknowledgments

This work was supported by the Claude D. Pepper Older Americans Independence Center grant # P30 AG24832 and R01 AG18311.

References


Key points

- Diabetic dyslipidemia includes elevated triglyceride levels, low levels of high-density lipoprotein cholesterol, and small cholesterol-depleted low-density lipoproteins.
- Insulin resistance is the main underlying abnormality causing diabetic dyslipidemia.
- The dyslipidemia is found in prediabetics and people with type 2 diabetes mellitus.

Introduction

Any discussion of lipid and lipoprotein metabolism in diabetes mellitus, especially type 2 diabetes mellitus (T2DM), must consider the role of insulin resistance (IR). IR plays a central role in the abnormal lipid and lipoprotein metabolism of T2DM [1] as evidenced by the characteristic set of lipid and lipoprotein abnormalities accompanying IR, even in the absence of frank hyperglycemia or abnormal glucose tolerance. Individuals with IR have low plasma levels of HDL-C and elevations in plasma triglyceride (TG) levels compared with levels seen in individuals who have normal insulin sensitivity. There is also an increase in the proportion of low-density lipoprotein (LDL) particles that are small, dense, and cholesterol ester-poor. Importantly, total cholesterol levels in insulin-resistant individuals are generally comparable to levels in people who are insulin sensitive, reinforcing the need for more detailed measures of lipoprotein to assess CVD risk [2–4]. Lipid derangements seen with IR progressively worsen across a continuum of declining glucose tolerance, from normal glucose tolerance (NGT) through impaired glucose tolerance (IGT) and T2DM [5]. Importantly, the relationship between IR and dyslipidemia extends across major ethnic groups, including African- and Hispanic Americans. Among women, IR and T2DM seem to exert a greater negative impact on several CVD risk factors, including TG and HDL-C levels and LDL particle size. This may, at least in part, account for the greater relative increase in coronary heart disease (CHD) risk observed in women with T2DM compared with their male counterparts [6].

IR and T2DM are also associated with increases in very low-density lipoprotein (VLDL) particle number and VLDL TG concentration, and a predominance of larger, more buoyant VLDL1 particles. In contrast, HDL particle number and size decrease in the face of increasing IR [2,3,5]. Plasma levels of apolipoprotein B100 (apo B100), the defining protein of the atherogenic lipoprotein series, which includes VLDL, intermediate-density lipoproteins (IDL), and LDL, are increased in the setting of IR [1]. This is particularly true when hypertriglyceridemia is also present, as is usually the case in individuals with IR or T2DM. In contrast, plasma levels of apolipoprotein A1 (apo A1), the surface protein unique to HDL particles, are reduced with IR [1].

The derangements in lipid and lipoprotein physiology discussed thus far occur in the context of the fasting state. Importantly, however, IR and T2DM are also associated with disordered postprandial lipid metabolism [7]. Although the severity of postprandial hyperlipidemia is typically closely related to the fasting plasma TG level [7], postprandial dyslipidemia has been demonstrated in people with T2DM even in the setting of normal fasting TG levels and optimal blood glucose control [8]. Although postprandial dyslipidemia has been associated with increased CVD prevalence in individuals without diabetes [7,9,10], this relationship has not been well characterized in people with IR or T2DM [7,11]. Importantly, a study in 150 individual with diabetes with or without coronary artery disease (CAD) demonstrated no differences in several aspects of postprandial lipid and lipoprotein levels [12]. This is probably indicative of the nearly universal abnormalities of postprandial lipid metabolism in T2DM. On the other hand,
several recent studies in large cohorts, most without diabetes, have demonstrated that nonfasting TG levels are as predictive or even more predictive of future CVD events than fasting TG concentrations [13,14].

**Lipoprotein metabolism in normal individuals and in type 2 diabetes mellitus**

**Key points:** Triglyceride levels in blood are regulated by the size and number of VLDL secreted from the liver and the efficiency with which triglycerides are cleared from the plasma by lipoprotein lipases. In T2DM, greater numbers of larger, more triglyceride-enriched VLDL are secreted from the liver and lipoprotein lipase activity may be modestly reduced; both lead to increased levels of triglyceride in the blood. HDL cholesterol levels are reduced for several reasons, including increased transfer of that cholesterol to very low- and low-density lipoprotein, in exchange for their triglyceride, by cholesterol ester transfer protein. LDL particles lose some of their cholesterol to VLDL via the same exchange mechanism, which generate smaller, denser, LDL.

Lipoproteins are macromolecular complexes consisting of core lipids, mainly TG and cholesterol esters (CEs), and surface phospholipids, free cholesterol, and one or more apolipoproteins. Five distinct major classes of lipoproteins have been characterized based on physical characteristics, including separation by ultracentrifugation, molecular weight, diameter, and chemical composition. Lipoprotein subclasses have been further defined based on subtle differences in physical and chemical properties. Notably, in addition to providing structural stability, the surface apolipoproteins impart critical functions to their respective lipoprotein particles [15].

Assembly and secretion of chylomicrons, which are the largest and most TG-enriched lipoprotein particles, occur in the postprandial state, when dietary TGs and CEs must be transferred from enterocytes of the small intestine into the circulation. The assembly and secretion of chylomicrons by the enterocyte is in several ways analogous to VLDL assembly and secretion by the hepatocyte: whereas apo B100 is the necessary, prototypic surface protein associated with VLDL, IDL and LDL, apo B48, a truncated form of apo B100, plays the same role for chylomicrons. Microsomal triglyceride transfer protein (MTP) packages apo B48 with core lipids and is essential to the formation of chylomicrons. Newly assembled chylomicrons, also carrying apolipoproteins A1 and A4, are secreted into the lymphatic system and eventually enter the venous circulation [15]. Studies in animal models of IR and/or diabetes have demonstrated increased intestinal secretion of apo B48-containing lipoproteins, accompanied by increased expression, mass and activity of intestinal MTP [16] and, under certain dietary conditions, an increase in intestinal de novo lipogenesis, likely mediated through activation of a major transcription factor, the sterol response element binding protein 1c (SREBP-1c) [16]. Moreover, studies of men with varying degrees of IR or with T2DM demonstrated increased intestinal production of apo B48-containing TG-rich lipoprotein (TRL) particles [17]. Increased intestinal expression of MTP in people with T2DM may also play a role here. Lewis and Adeli and their colleagues have furthered our knowledge of the assembly and secretion of apo B48-containing lipoproteins, including demonstration of inhibition of chylomicron secretion by acute administration of insulin and glucagon-like peptide 1 [18]. The conflicting effects of acute insulin and chronic hyperinsulinemia in states of IR suggest that the latter also develops in the intestine.

After entering the plasma, chylomicrons acquire apolipoproteins C1, C2 and C3 (apo C1, apo C2 and apo C3) from the surface of HDL particles. Chylomicron clearance is largely mediated through hydrolysis of core TGs by adipose tissue-derived lipoprotein lipase (LpL) in a process that requires apo C2 as an activator of LpL. In contrast, apo C3 can inhibit LpL-mediated lipolysis. The activity of other factors, including apo A5, which appears to facilitate LpL-mediated lipolysis [19], and angiopoietins 3 and 4 [20,21], which inhibit lipolysis, may also regulate postprandial lipid metabolism. The breakdown of chylomicron TG releases fatty acids (FAs), which are taken up by the local tissue, and produces chylomicron remnant particles. These remnants acquire apolipoprotein E (apo E) from HDL particles and are also enriched in CEs derived from exchange of core lipids with LDL and HDL, the latter process being mediated by cholesteryl ester transfer protein (CETP).

Ultimately, chylomicron remnant particles are cleared mainly by hepatic uptake through the interactions of apo E with the hepatic LDL receptor, the LDL receptor-related protein (LRP), and/or cell-surface heparan sulfate proteoglycans. Additionally, hepatic lipase (HL) further hydrolyzes chylomicron remnant particle TG and may augment uptake by the liver. Apo C1 and C3 modulate chylomicron clearance by the liver by interfering in the binding of apo E to its receptors [15].

IR and T2DM affect chylomicron and remnant metabolism negatively at several points in the schema described earlier. For example, there are modest decreases in LpL activity and increases in apo C3 relative to apo C2 levels. Increased apo C3 secretion into plasma has been demonstrated in patients with IR and hypertriglyceridemia [22,23], with impaired LpL-mediated lipolysis of chylomicron TG as well as decreased hepatic uptake of remnant particles the likely results. Recent studies with antisense to apo C3 in hypertriglyceridemia support the important role of this apoprotein in metabolism of TG-rich lipoproteins [24]. Clearance of chylomicrons is competitive with clearance of VLDL and so hepatic overproduction of TG-rich VLDL particles in T2DM can, itself, lead to reduced efficiency of chylomicron clearance [1]. This results in increases in circulating TRLs, which are known to be particularly atherogenic [25].

In the fasting state, hepatic VLDL assembly and secretion predominate. Utilization of nascent apo B100 for the assembly
and secretion of VLDL is, to a large extent, limited by substrate availability, namely TG, as apo B100 is constitutively synthesized [26]. In the presence of adequate TG, apo B100 is progressively lipitated through the actions of hepatic MTP, and the newly assembled VLDL particle is transported to the Golgi apparatus, where it may undergo further lipidation before secretion into plasma. The size of the nascent VLDL particles is also determined by TG availability, with a predominance of larger, more buoyant, more TG-enriched VLDL1 and fewer smaller, denser VLDL2 particles produced in the presence of excess hepatocyte TG. Decreased availability of core lipids and/or insufficient MTP activity can lead to both co- and posttranslational degradation of apo B100. Insulin can acutely target apo B100 for post-translational degradation and this may be important in the postprandial period. However, this effect may be significantly diminished in the presence of IR and concomitant chronic hyperinsulinemia and steatosis [27]. Thus, a combination of IR and excess hepatic TG availability results in increased hepatic secretion of both apo B100 and TG even when hyperinsulinemia is present [28–30]. Further complicating this scheme is the ER stress that develops in the presence of IR and fatty liver [31]; apo B100 degradation may be increased by ER stress, leading to less hepatic VLDL secretion and greater steatosis [32]. Thus, a combination of several factors common to IR and T2DM, including hyperinsulinemia, increased FAs, increased lipogenesis, steatosis, ER stress, and apoB degradation, will all play important roles in determining the quantity and size of VLDL secreted from the liver [33].

TG availability is a major driver of VLDL assembly and secretion. FA contribution to the hepatic TG pool is, therefore, a key regulator of that process. Circulating FAs from the peripheral tissues, notably from adipose tissue, FAs derived from chylomicron and VLDL remnant uptake by the liver, and FAs obtained through de novo hepatic lipogenesis are the three sources of hepatic FA for TG synthesis; the major source, by far, being circulating FA [34]. In states of IR and T2DM, all three sources increase hepatic FA [35]. FA flux to the liver is increased as a result of IR in adipocytes, and this can directly stimulate hepatic secretion of apo B-containing lipoproteins [36,37]. Elevated levels of apo B48- and apo B100-containing remnant particles provide TG for hepatic uptake [28]. De novo lipogenesis, the third source of hepatic TG, is likely upregulated via insulin-mediated stimulation of SREBP-1c, a major lipogenic transcription factor. Recent studies indicate that mTOR signaling, both via the mTORC1 and mTORC2 complexes, is central to the “selective” insulin response in the lipogenic pathway in an otherwise insulin-resistant liver [38]. If there is concomitant hyperglycemia, there will also be induction of another transcription factor, carbohydrate responsive element binding protein (ChREBP), which also activates genes required for lipogenesis [30]. Interestingly, the hepatic expression of another nuclear transcription factor, peroxisome proliferator-activated receptor γ (PPARγ), is increased in a number of mouse models of IR and dyslipidemia [39], as well as in humans with hepatic steatosis. In humans, nonalcoholic fatty liver disease (NAFLD), which is closely associated with IR, is characterized by increased hepatic TG derived from de novo lipogenesis [34].

The fate of VLDL particles after secretion initially parallels that of chylomicrons; hydrolysis of VLDL TG by LpL, modulated by the relative proportions of apo C2 and apo C3. Additional regulation of LpL activity derives from the interactions of apo AV, which may facilitate the interaction of TG-rich lipoproteins with LpL at the endothelial cell surface [40] and the angiopoietin-like proteins (angptl3 and angptl4), which interfere with the activity of LpL [41]. Lipolysis of core TG yields smaller, denser VLDL remnants usually referred to as IDL; the efficiency of this process is impaired in the setting of T2DM. Hepatic uptake of VLDL remnants is similar to that of chylomicron remnants, except that VLDL remnants or IDL may undergo further metabolic modifications to generate LDL particles. HL is thought to play an important role in this process. However, despite the increased levels of HL observed with IR and T2DM, conversion of VLDL to LDL is typically reduced. LDL particles are primarily composed of core CEs associated with surface apo B100. As with VLDL remnants and IDL, clearance of LDL occurs mainly through hepatic LDL-receptor-mediated uptake. IR and poorly controlled T2DM can be associated with a reduction in LDL receptors, thereby limiting removal of these particles from the circulation [15,28].

LDL in the circulation can participate in CETP-mediated exchange of its core CE for VLDL or chylomicron TG, resulting in a TG-enriched, CE-depleted LDL particle. Subsequent lipolysis of the TG-enriched LDL particle by LpL or HL produces small, dense LDL. The characteristic hypertriglyceridemia of T2DM is associated, therefore, with the presence of small, dense LDL. It has also been proposed that the large VLDL3 particles that predominate in the setting of T2DM hypertriglyceridemia are avid acceptors of CETP transferred CEs from HDL and preferentially undergo catabolism to give rise to small, dense LDL [15,42]. In addition, the increased HL activity observed in T2DM could play an important role in the production of small, dense LDL particles [15]. The role of small dense LDL has been a longstanding topic of debate, particularly as to whether they are more atherogenic than other apo B-containing lipoproteins [43,44].

The “anti-atherogenic” HDL differ considerably in structure and function from the apo B-containing lipoproteins. HDL begins as cholesterol-poor phospholipid discs containing surface apo A1 that is secreted by the liver and intestine. Transfer of intracellular or plasma membrane cholesterol to these nascent particles occurs via the transport protein, ATP-binding cassette transporter A1 (ABCA1). Studies by Parks and colleague demonstrated the critical role of hepatic ABCA1 in the maintenance of plasma HDL cholesterol concentration [45], but transfer of cholesterol from peripheral cells also contributes to HDL. This transfer is coupled with cholesterol esterification by the enzyme lecithin:cholesterol acyltransferase (LCAT), accompanied by movement of the product CEs to the core of the maturing HDL3 particle. Apo A1 mediates this process through its ability to activate LCAT. Repeated cycles
of this process, together with transfer of cell cholesterol to the maturing HDL particle via ATP-binding cassette transporter G1 (ABCG1) and scavenger receptor B1 (SRB1), give rise to mature, CE-rich HDL2. Mature HDL particles can deliver both free and esterified cholesterol to the liver via interaction with SRB1 [1]. There may be other pathways for HDL uptake by the liver as well. The process just described, at least when the source of free cholesterol is a peripheral cell, has been called reverse cholesterol transport (RCT). It is believed, based on mainly mouse studies, the RCT from cholesterol-laden macrophages in the artery plaque, with eventual hepatic uptake and metabolism, is a critical function of HDL. The fate of apo A1 is less certain, but it is known that significant quantities are taken up and degraded by the kidneys and liver. If CETP-mediated exchange of core lipids is increased, as is the case in the dyslipidemia of T2DM, apo A1 dissociates from the smaller, triglyceride-rich HDL, and clearance of “free” apo A1 by the kidney follows, reducing the plasma apo A1 levels and HDL particle number [15,46].

**Therapeutic approaches to the dyslipidemia of insulin resistance and type 2 diabetes mellitus**

**Key points:** The rationale for treating the diabetic dyslipidemia is clear, it is associated with significantly increased risk of cardiovascular events. Weight loss, exercise, and diet modifications are key to successful treatment of all of the lipid abnormalities in patients with T2DM. Statins are the mainstay of therapy—they lower LDL cholesterol up to 50%, lower triglycerides about 15–25%, and have minimal effects of HDL cholesterol levels. Statins significantly lower cardiovascular events in people with diabetes. Benefits of treatment of high triglycerides and low levels of HDL cholesterol after reaching optimal treatment status with statins is unclear.

**Rationale for treating LDL**

Over the past 50 years, results from epidemiologic studies, animal models, and clinical trials have clearly demonstrated that elevated LDL-C is a major risk factor for CV morbidity and mortality [47]. In addition, it has been demonstrated that aggressively lowering LDL-C levels can decrease CV risk [48]. As a result, a system of guidelines for LDL goals for therapy developed over the past 25 years; a quick review of the progression of these guidelines was published recently [49].

Beginning in 2004, there had been a push to drive LDL-C goals lower and lower with the hope that this would lead to an improvement in CV outcomes. In fact the NCEP issued a “white paper” suggesting an even lower LDL-C goal of <70 mg dL⁻¹ for patients at very high CV risk [50], and this was reiterated in the 2006 AHA/ACC guideline paper [51]. These reports were based in part on data from the Treating to New Targets (TNT) study, which demonstrated that intensive therapy with atorvastatin 80 mg compared with atorvastatin 10 mg significantly reduced the rate of major CV events by 22%. In TNT, the end-of-treatment mean LDL-C levels were 98.6 mg dL⁻¹ and 77 mg dL⁻¹ with atorvastatin 10 mg and 80 mg, respectively [52]. The results of the Pravastatin or Atorvastatin Evaluation and Infection Therapy—Thrombolysis in Myocardial Infarction 22 (PROVE-IT—TIMI 22) study [53], in which patients with acute coronary syndrome who had their LDL-C levels reduced to 62 mg dL⁻¹ fared better than those patients whose LDL-C was 95 mg dL⁻¹, support the TNT results.

Statin trials have also been conducted specifically in patients with T2DM. In the Collaborative Atorvastatin Diabetes Study (CARDS) trial, treatment with atorvastatin resulted in a significant reduction in major CV events irrespective of pretreatment cholesterol levels [54]. On the other hand, the Atorvastatin Study for Prevention of Coronary Heart Disease Endpoints in Non-Insulin-Dependent Diabetes Mellitus (ASPEN), which randomized patients with T2DM to receive 10 mg of atorvastatin or placebo in a 4-year study, did not demonstrate a significant reduction in the primary composite endpoint of CV death with atorvastatin treatment [55]. Several post hoc subgroup analyses of large trials not specifically targeted to diabetics did show similar benefit in the latter groups; they will be reviewed further in this chapter.

The very recent AHA/ACC guidelines [56] suggested, however, major changes in the way healthcare professionals should lower blood cholesterol levels to reduce cardiovascular disease. A strong point was the change in the use of initial LDL-C levels as the approach to initiation of treatment to identification of four groups of individuals whose risk was high enough to be considered for statin therapy: those with LDL-C >190 mg dL⁻¹; those with prior CVD; those with diabetes and LDL-C ≥70 <190 mg dL⁻¹; those without CVD or diabetes but an LDL-C ≥70 ≤190 mg dL⁻¹ who have a 10-year risk for CVD ≥7.5%. The new guidelines do not use MetS as a criteria for statin therapy. Clearly, the presence of diabetes is a major factor in these new guidelines. The new guidelines then declared that there were no data from randomized clinical trials (RCTs) to support the use of LDL-C goals to determine the intensity of treatment. The guidelines panel concluded that use of statins was supported strongly by RCT results and that high-intensity treatment was appropriate in most cases, particularly in patients with diabetes. However, the panel found no evidence for the use of non-statin drugs as “add-on” treatment to statins [56]. The new guidelines have, as may be expected, generated significant controversy. In particular, the conclusion that there is no evidence that “lower is better” is thought by some to fly in the face of the trials noted above in this section as well as the meta-analyses of the statin trials [48], epidemiologic data, animal-based studies, and natural genetic examples.

**Rationale for treating HDL and TG**

While LDL-C remains a primary target of cholesterol-modifying therapy for the primary and secondary prevention of CVD in the general population and among people with DM, HDL-C and TG have been given high importance, especially in higher risk
individuals [25,47]. The characteristically elevated TG levels and decreased HDL-C with IR, in the presence or absence of T2DM [2,3,5] led to the recommendation that non-HDL-C (total cholesterol - HDL-C) be considered a surrogate marker for the apo B100-containing lipoproteins [57,58]. The new AHA/ACC guidelines [56] concluded, however, that there is no evidence to target either LDL-C or non-HDL-C. The new guidelines also stated that other targets, such as TG or HDL-C have no support based on RCT; in fact, data from two trials where niacin was added to statins, Aim High [59] and HPS2-THRIVE (not yet published), and one in which fenofibrate was added to statin therapy, ACCORD [60], indicate no overall benefit compared to patients treated with statins alone. The ADA on the other hand, emphasizes the prime importance of attaining target LDL-C levels, but employs a different approach to the management of TGs and HDL-C (Table 17.1). The ADA guidelines recommend lowering TG levels to <150 mg dL^{-1} and include gender-specific HDL-C goals; namely, >40 mg dL^{-1} in men and >50 mg dL^{-1} in women [61–63].

With the AHA/ACC and ADA guidelines at odds, it is valuable to review the rationale for treatment to improve HDL-C levels in people with IR or T2DM. First, there is evidence from numerous epidemiologic studies in which low HDL-C is an independent predictor of CHD morbidity and mortality, including studies involving large numbers of people with IR or T2DM [64]. Second, evidence from prospective intervention trials and secondary analyses has suggested a benefit of raising HDL-C levels on the risk of CHD events, independent of changes in other lipid and nonlipid risk factors [65]. Of note, a post hoc analysis of the Bezafibrate Infarction Prevention (BIP) Study focusing on the subgroup of participants with the MetS revealed a reduced incidence in MI associated with significant improvements in HDL-C and TG levels [66]. Furthermore, in that analysis a cardiac mortality benefit was demonstrated among subjects having “augmented” (i.e., four or five) features of the MetS, as defined by the ATP III. In VA-HIT, where increases in HDL-C were predictive of reductions in CV endpoints in the group treated with gemfibrozil [65], the greatest overall benefit of treatment was observed in subgroups with IR, with or without T2DM [67]. On the other hand, the recent negative niacin trials, Aim High [59] and HPS2-THRIVE, do not support the use of adding niacin to statin therapy. A subgroup, post hoc analysis of Aim High did suggest, nonsignificantly, that participants with high TG and low HDL-C at baseline did benefit [68].

Data regarding the independent effects of plasma TG levels on CHD risk are even more controversial, partly due to the close interrelationships between levels of TGs and the atherogenic apo B-containing lipoproteins. TG levels have nevertheless been shown to predict CHD risk in observational studies [69]. However, in the largest cohort study to date, Danesh and colleagues found that while TG was a strong predictor of CVD outcomes in more than 300,000 people, it lost all predictive power after adjustments for both HDL and non-HDL cholesterol [70]. One could argue, however, that adjustment for non-HDL, which contains cholesterol carried in TG-rich lipoprotein, is not a valid approach. In several clinical trials, stratification of individuals by TG and HDL-C levels has revealed higher risk subgroups with dyslipidemia that have benefited most from LDL-C-lowering interventions with respect to CHD risk reduction [71]. The FIELD Trial, in which 10,000 diabetics were randomized to placebo or fenofibrate, did not show an overall CVD benefit for the fenofibrate-treated group. However, a subgroup with baseline TG over 200 mg dL^{-1} and HDL-C less than 40 mg dL^{-1} showed a 30% reduction in the combined primary endpoint of CAD death and nonfatal myocardial infarction [72]. Similarly, in the ACCORD Lipid Trial, which was negative overall, 17% of the subjects with baseline TG in the upper tertile (>200 mg dL^{-1}) and HDL-C in the lower tertile (<34 mg dL^{-1}) had a 28% reduction in the primary endpoint of nonfatal myocardial infarction, nonfatal stroke, and CVD death [60]. On the other hand, no clinical trials thus far have demonstrated an independent relationship between TG-lowering and CHD risk reduction, including VA-HIT, where TG levels fell 31%.

As mentioned earlier, it is worth noting that postprandial lipemia has been associated with CAD in individuals without DM, whereas the evidence is less convincing or absent in individuals with IR or DM [7]. Data from two cross-sectional studies do not support an association between postprandial lipemia and the presence of CAD among people with T2DM [12]. Another small study suggests that postprandial numbers of

Table 17.1 American Diabetes Association dyslipidemia treatment goals and recommendations for adults with diabetes [62]

<table>
<thead>
<tr>
<th>Treatment goals</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LDL-C</strong></td>
</tr>
<tr>
<td>In patients without overt CVD:</td>
</tr>
<tr>
<td>Primary goal is LDL-C &lt;100 mg dL^{-1}</td>
</tr>
<tr>
<td>Decrease LDL-C by 30–40% in patients &gt;40 years of age regardless of baseline LDL-C</td>
</tr>
<tr>
<td>In patients with overt CVD:</td>
</tr>
<tr>
<td>Treat with statin to decrease LDL-C by 30–40%</td>
</tr>
<tr>
<td>Option: treat with high-dose statin to reduce LDL-C to &lt;70 mg dL^{-1}</td>
</tr>
<tr>
<td>Decrease TG to &lt;150 mg dL^{-1}</td>
</tr>
<tr>
<td>Increase HDL-C to &gt;40 mg dL^{-1}; consider increasing HDL-C to &gt;50 mg dL^{-1} in women</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Treatment recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Medical nutrition therapy</strong></td>
</tr>
<tr>
<td>Weight loss in overweight and obese individuals</td>
</tr>
<tr>
<td>Diet with reduced intake of total fat (~30% of total calories), saturated fat (~7% of total calories), and trans fatty acids (minimal); cholesterol &lt;200 mg day^{-1}</td>
</tr>
<tr>
<td>Increased exercise</td>
</tr>
<tr>
<td>Smoking cessation</td>
</tr>
<tr>
<td><strong>Pharmacologic therapy</strong></td>
</tr>
<tr>
<td>Statin therapy to decrease LDL-C</td>
</tr>
<tr>
<td>Fibrate therapy to decrease TG and increase HDL-C</td>
</tr>
<tr>
<td>Combination therapy with statin and another agent if necessary to achieve lipid targets</td>
</tr>
</tbody>
</table>

Source: Buse 2007 [62].
small remnant particles may contribute to the severity of angiographic CAD in T2DM [73]. Larger prospective studies are needed before recommendations can be made regarding the potential utility of postprandial hyperlipidemia as a predictor of CV risk in IR or DM. This, along with the lack of simple, clinically useful uniform measures of postprandial lipemia, explains the lack of guidelines for therapy of postprandial hyperlipidemia.

**Treatment options**

**Diet and lifestyle modification**

Appropriate lifestyle interventions involving diet and exercise can lead to weight loss, and since obesity is a major contributor to the IR and T2DM, weight loss should provide significant benefits to individuals with these problems. The Diabetes Prevention Program [74] demonstrated that a 6% weight loss and the addition of two hours of exercise per week was associated with a 58% reduction in the incidence of new T2DM in a group with IGT. Similar results were obtained in the Finnish Diabetes Prevention Study using similar interventions [75]. The 1-year results of Look AHEAD (Action for Health in Diabetes), a study of weight loss and diet in overweight and obese individuals, reported that intensive lifestyle intervention resulted in clinically significant weight loss in people with T2DM [76]. Participants in the intensive lifestyle intervention group achieved an average loss of 8.6% of initial body weight and a 21% improvement in CV fitness. These participants had a significantly greater decrease in the number of medicines used to treat their DM and blood pressure as well as a greater improvement in glycemic control when compared with participants receiving only diabetes support and education. These results, for the most part, withstood the demands of a 10-year follow-up, although the primary endpoint, decreased mortality in the intensive intervention group, was not reached [77].

In addition to simply reducing caloric intake, interventions that alter the nutrient composition of the diet may have benefits on risk factors for CVD and T2DM. There has been much debate, however, both in the popular press and in the medical community, regarding which diet strategy is most effective for both weight loss and reductions in CV risk. Conclusions that may be drawn from these diet interventions must be considered with the following caveat: We do not know if the changes that we see in plasma TG levels and, in particular, in HDL-C concentrations, are indicative of changes in risk for CVD. We are, in fact, extrapolating from large epidemiologic databases that show that increased TG and reduced HDL-C levels are associated with increased CV risk; this extrapolation remains to be proven in a clinical trial with dietary intervention.

**Pharmacologic therapy**

Before discussing in detail specific pharmacologic therapies, it is important to note that improving glycemic control in T2DM can result in mild to modest improvement in the dyslipidemia of diabetes. This is particularly true for hypertriglyceridermia in the setting of poor glycemic control; furthermore, the degree of TG-lowering may be related to the magnitude of improvement in glycemia. Concomitant with optimization of glucose control, therapeutic lifestyle changes (TLC), including dietary modification, weight reduction, and regular physical activity, should be initiated prior to, or simultaneously with, pharmacologic therapy. Pharmacologic therapy for dyslipidemia should be considered in the context of these changes [61,63,78]. Additionally, it must be noted that the new AHA/ACC guidelines for the treatment of cholesterol support the use of statins in several high-risk groups but do not support additional medications in combination with statins [56]. The new guidelines do allow non-statin drugs for patients intolerant to statins and for individuals with initial LDL-C levels greater than 190 mg dL⁻¹ (usually with familial hypercholesterolemia). It is noteworthy that the ADA/ACC guidelines for the management of lipoproteins in patients with cardiometabolic risk (which includes people with diabetes), while supporting the addition of non-statin drugs for people not at LDL-C goal on statin monotherapy, provided little support for non-statin drugs for the treatment of TG and HDL-C abnormalities [63].

**3-hydroxy-3-methylglutaryl coenzyme A (HMG CoA) reductase inhibitors**

The HMG CoA reductase inhibitors, or “statins,” are the most effective pharmacologic agents for reducing LDL-C concentrations and should be considered first-line therapy for this purpose, both in individuals with and without T2DM [48,63]. Several agents are available for clinical use, including lovastatin, fluvastatin, pravastatin, simvastatin, atorvastatin, rosuvastatin, and pitavastatin. As a class, they produce LDL-C reductions of 18–55%, with more modest changes in HDL-C and TG levels (5–15% elevations and 7–30% reductions, respectively) [79]. The new AHA/ACC guidelines suggest high-intensity statin therapy for most individuals with diabetes.

The effectiveness of statin therapy in primary and secondary CV risk reduction in both the general population with hypercholesterolemia and in people with diabetes is extremely strong (Table 17.2) [48,62]. Included in the meta-analyses were data from the following trials: The CARDS, a primary prevention study of 2838 participants with T2DM, demonstrated a statistically significant 37% reduction in major CV events among participants on low-dose atorvastatin treatment [54]. The Medical Research Council/British Heart Foundation Heart Protection Study (MRC/BHF HPS) included a large cohort of participants with T2DM, approximately half of whom had no evidence of existing coronary or other occlusive arterial disease [80]. Simvastatin therapy produced substantial reductions in the risk of coronary events, stroke and revascularizations in participants with DM, irrespective of pre-existing occlusive arterial disease or lipid concentrations. Further support was added by a subgroup analysis of the Long-Term Intervention with Pravastatin in Ischemic Disease (LIPID) trial, which demonstrated a reduction in CV events, including stroke,
in participants with IFG or T2DM and established CHD [81]. Moreover, the absolute risk reductions seen in those with IFG or T2DM were greater than those seen in the overall study population. These findings were supported and extended by the results of post hoc analyses of participants with IFG or DM in the Scandinavian Simvastatin Survival Study (4S) [82]. Another post hoc analysis demonstrated an increased risk of CHD events and greater benefit with simvastatin therapy among 4S participants with low HDL-C and elevated TG accompanying elevated LDL-C levels at baseline, compared with participants with isolated elevations in pretreatment LDL-C [71]. The Cholesterol and Recurrent Events (CARE) trial demonstrated a beneficial effect of pravastatin treatment on coronary outcomes in the overall study population as well as in the subgroups with DM or IFG, the latter being defined in that study by fasting glucose levels between 110 to 125 mgdL$^{-1}$ [83]. Finally, a post hoc analysis of the TNT study that examined those participants who met the ATP III criteria for the MetS, but excluded individuals with T2DM, suggested an incremental benefit of high-dose atorvastatin therapy in this group [84].

The ASPEN trial, which included primary and secondary prevention participants, did not, however, demonstrate a significant reduction in the primary composite endpoint with atorvastatin [55]. The negative finding was attributed to certain features of the study design, as well as to changes in lipid treatment guidelines during the course of the study that necessitated protocol changes. Similarly, statin therapy did not impart a benefit with regard to coronary events in the subgroup of participants with T2DM in the Anglo-Scandinavian Cardiac Outcomes Trial—Lipid-Lowering Arm (ASCOT-LLA) [85]. The lack of benefit was ascribed to a low coronary event rate, a result stemming in part from the early cessation of the trial because of a marked benefit of statin therapy in the general study population. In the nonblinded Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT-LLT), the lack of effect with pravastatin therapy in participants with T2DM is believed to be due to the high usage of lipid-lowering therapy in the “usual care” control group, resulting in a smaller differential in LDL-C levels between the statin and control groups compared with that seen in other large statin trials [86].

A recent issue that must be addressed is how to balance aggressive statin treatment with recent data indicating an increase in risk for new diabetes in individuals receiving statins in large clinical trials. Although an early report from the WOSCOPS suggested that statins may prevent incident diabetes, JUPITER found the opposite, with a rate of incident diabetes of 30% in the rosvastatin group versus 2% in the placebo group [87]. Several meta-analyses since then have indicated that statin treatment is associated with about a 9% increase in new diabetes [88] and that this risk is greater in people receiving higher doses of statins [89] as well as in people with more criteria for the MetS [90]. The meta-analyses presented to date do not suggest heterogeneity in the effects of different statins on incident diabetes, although some small, short-term studies indicate that certain statins may be less likely to raise blood glucose or HbA1c levels. In any event, it is clear that the benefit of statins in lowering CVD events far outweighs the risk of new onset T2DM in high risk and secondary prevention patients. Obviously, a risk for new onset diabetes is not relevant in people with diabetes, but health professionals should consider these new data because they indicate that control of existing diabetes is likely to worsen modestly due to long-term statin therapy.

Many “lipid/diabetes experts” believe that the greatest potential value of statin therapy in the treatment of the dyslipidemia of T2DM would be in combination with other lipid-modifying agents. This would clearly be the case when LDL-C-lowering to levels below 70 mgdL$^{-1}$ is considered an appropriate goal,
for example, in patients with T2DM who have CVD or multiple poorly controlled risk factors [50]. This view is now dependent on whether one accepts the newest AHA/ACC guidelines [56] or the ADA/ACC guidelines [63]. For those following the ADA recommendations, LDL-C below 79 mg dL\(^{-1}\) may be attained through up-titration of statin doses or through combination with other agents, such as plant stanol and sterol esters, bile acid-binding resins and inhibitors of enteral cholesterol absorption. However, neither guideline give much support for the addition of niacin or peroxisome proliferator-activated receptor \(\alpha\) (PPAR\(\alpha\)) agonists to statins.

**Nicotinic acid**

Nicotinic acid (niacin) has been used for more than 50 years to lower LDL-C and TGs, and increase the level of HDL-C. Indeed, niacin is the most potent HDL-C-raising pharmacologic agent available. Of note, we still, at this late time in the life of niacin, do not have definitive evidence regarding its mechanism(s) at the molecular level [91].

Niacin is available in short-acting and extended-release forms and is typically associated with increases in HDL-C of 15–35%, with 20–35% reductions in TG levels [92]. In contrast, it has a more modest effect on LDL-C levels, with reductions in the range of 10–20%. For this reason, it is not recommended as first-line therapy for LDL-C-lowering. It is, however, suggested for use as monotherapy in individuals with the diabetic lipid phenotype but without concomitant elevations of LDL-C. For people who exhibit elevated LDL-C in addition to abnormal HDL-C and TG levels, niacin can be used in combination with statins, or other LDL-C-lowering agents. Niacin can also reduce levels of lipoprotein (a) between 15–25%.

Despite the array of potentially beneficial effects on plasma lipids, the role of niacin has now been significantly diminished as an agent to prevent CVD. As described earlier, two recent completely negative trials where niacin was added to statin, far outweigh the positive outcome in the Coronary Drug Project (CDP) where niacin was given as monotherapy. In addition, niacin is associated with glucose intolerance and hyperglycemia, particularly at higher doses. The basis for these adverse effects on carbohydrate metabolism appears to result from niacin-induced IR [93]. Increased incident diabetes and worsening of existing diabetes was evident in HPS2-THRIVE [94].

At the present time, niacin treatment should be limited to individuals with familial hypercholesterolemia who do not reach goals on statins and other agents and possibly in statin-intolerant patients at low risk for diabetes. Physicians wishing to use niacin in other situations, that is, in patients with very low levels of HDL-C or high levels of Lp (a), must consider the absence of evidence for benefit.

**Bile acid sequestrants (BAS)**

This is another class of agents that have long been used in the management of lipid disorders. They have also been successful in monotherapy RCTs for preventing cardiovascular events [95]. In the United States, currently available drugs in this category include cholestyramine, colestipol and colesevelam. BAS lower LDL-C by reducing the enterohepatic return of cholesterol to the liver, leading to increased use of hepatic cholesterol for the bile acid synthesis pathway and consequent lowering of hepatic cholesterol content. The latter leads to upregulation of hepatic LDL receptors, which in turn lowers LDL-C levels. Typical improvements seen with BAS therapy include 15–30% reductions in LDL-C and smaller increases in HDL-C, in the order of 3–5%. In some cases, therapy with agents in this class increases hepatic VLDL production, thereby increasing plasma TG levels and worsening hypertriglyceridemia. These drugs should therefore be avoided in individuals with severe hypertriglyceridemia (TG > 400 mg dL\(^{-1}\)), and caution should be exercised when considering bile acid sequestrants for LDL-C-lowering in patients with moderately elevated TG levels (TG > 200 mg dL\(^{-1}\)), a group that encompasses a considerable number of people with T2DM. In addition, BAS should be avoided in patients with DM complicated by autonomic neuropathy and constipation, due to the gastrointestinal side effects associated with these drugs, particularly cholestyramine and colestipol. Use of ezetimibe as an adjunct to statin treatment would depend on whether one believes in the AHA/ACC or the ADA/ACC guidelines for lowering LDL-C.

Over the past several years, there has been a renewed interest in BAS, especially as a potential LDL-C-lowering agent in people with diabetes. Preclinical studies of the role of farnesoid X receptor (FXR) in the regulation of carbohydrate metabolism have raised interest in the glucose-lowering effects of BAS [96]. Of note, glucose levels improved and there were reductions in HbA1c levels in the range of 0.4% in a recent study with colesevelam [97]. Indeed, colesevelam received an indication from the FDA for glucose lowering.

**Cholesterol absorption inhibitors**

Ezetimibe was the first available agent in a new class of molecules that interferes with cholesterol absorption by selectively inhibiting the uptake of cholesterol from the intestinal lumen at the level of the brush border of the enterocyte; for a number of reasons, it remains the only available agent in this class. A multicenter, randomized, double-blind, placebo-controlled trial demonstrated that ezetimibe 10 mg decreased LDL-C by 17–20% compared with placebo [98]. It has small effects on TG and HDL-C levels; these are significant in some, but not all, studies. Studies in individuals without DM have demonstrated that combined therapy with ezetimibe and a statin produces greater reductions in LDL-C and TG levels, and greater increases in HDL-C, than either therapy alone [99,100]. Several studies have specifically examined the effects of ezetimibe when added to statin therapy in individuals with the MetS and/or T2DM; the results were similar to those observed in groups without DM. Ezetimibe has not been demonstrated to reduce atherosclerosis in combination with a statin. More recently, in the Study of Heart and Renal Protection (SHARP) trial,
ezetimibe plus simvastatin significantly reduced CVD events in a population with moderate to severe renal failure; treatment did not, however, reduce progression to dialysis [101]. Use of ezetimibe as an adjunct to statin treatment would depend on whether one believed in the AHA/ACC or the ADA/ACC guidelines for lowering LDL-C.

**PPARα agonists**
The fibrates are a class of drugs that act as ligands for PPARα receptors. The latter are proteins that, as heterodimers with another protein, retinoid X receptor (RXR), bind to specific DNA sequences in the promoters of genes and activate them. The natural ligands for PPARα are likely to be FAs or their derivatives. PPARα are mainly expressed in the liver and activate genes important for hepatic lipid and lipoprotein metabolism [102]. Two fibrates, gemfibrozil and fenofibrate, are available in the United States. Typically, TG levels are reduced by 35–50% and HDL-C levels are increased by 5–20% with PPARα treatment; greater changes are seen with more extreme baseline abnormalities [103,104]. Three major fibrate monotherapy studies with CV events as the major outcome have included some participants with the MetS and/or T2DM. The Helsinki Heart Study (HHS) was a primary prevention study of hypercholesterolemic subjects in which gemfibrozil treatment reduced CV events by 35% in the overall group of participants [105] and had a similar benefit in the very small group of participants with T2DM [106]. In the VA-HIT trial, a secondary prevention study of men with "normal" LDL-C levels (112 mg dL⁻¹), moderately elevated TG levels (160 mg dL⁻¹), and very low levels of HDL-C (32 mg dL⁻¹), gemfibrozil treatment was associated with a 22% reduction in MI or CHD death in the overall study group [107] and a similar benefit in the 25% of participants with T2DM [108]. Although these studies support the use of gemfibrozil monotherapy therapy, particularly in subjects with T2DM, its use is limited because most of those individuals will already be treated with statins, and gemfibrozil increases the risk of statin-associated myositis when the two are used in combination. Unfortunately, several additional fibrate trials have failed to reach their primary endpoints. The Fenofibrate Intervention and Event Lowering in Diabetes (FIELD) Study was a monotherapy study in individuals with T2DM that enrolled both primary and secondary prevention participants [109]. Fenofibrate did not significantly reduce the risk of the primary outcome of coronary events (CHD death and nonfatal MI) compared to placebo; nonfatal MI was reduced, but CHD death was unchanged. A potential confounder of the FIELD study was an imbalanced statin "drop-in;" by the end of the study, nearly 40% of subjects in the placebo group had been exposed to statins, while only 20% of the fenofibrate group were treated with statins at some point in the study. Additional disappointment concerning the role of PPARα agonists derive from the overall results of the ACCORD Lipid Trial where fenofibrate therapy did not significantly reduce the primary CVD outcome when added to simvastatin treatment [60]. Thus, despite a subgroup analysis suggesting a dramatic (28%) benefit in 17% of the patients with baseline TG levels in the upper tertile (>204 mg dL⁻¹) and the lower tertile of HDL-C levels (<34 mg dL⁻¹), which parallels similar subgroup findings in fibrate monotherapy trials, a study specifically targeting such individuals will have to be successful before recommendations for combination therapy with fibrate and statin can be recommended. Very recently, aliglitazar, a dual PPARα/γ agonist, also failed to reduce CVD events in a population with T2DM [110]. At this time, therefore, PPAR agonists do not have evidence-based roles in the prevention of CVD in T2DM. At the time this chapter was being written, the VA Hospital System was planning to perform an RCT based on ACCORD, with all participants having TG >200 and HDL <35 mg dL⁻¹ while on statin monotherapy.

**References**


22 Cohn JS, Patterson BW, Uffelman KD, et al.: Rate of production of plasma and very-low-density lipoprotein (VLDL) apolipoprotein C-III is strongly related to the concentration and level of production of VLDL triglyceride in male subjects with different body weights and levels of insulin sensitivity. *Journal of Clinical Endocrinology & Metabolism* 2004;89:3949–3955.


41 Robciuc MR, Maranghi M, Lahikainen A, et al.: Angptl3 deficiency is associated with increased insulin sensitivity, lipoprotein lipase


57 Grundy SM: Low-density lipoprotein, non-high density lipoprotein, and apolipoprotein B as targets of lipid-lowering therapy. Circulation 2002;106:2526–2529.


71 Ballantyne CM, Olsson AG, Cook TJ, et al.: Influence of low high-density lipoprotein cholesterol and elevated triglyceride on
coronary heart disease events and response to simvastatin therapy in 4S. *Circulation* 2001;104:3046–3051.


86 The ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group: Major outcomes in moderately hypercholesterolemic, hypertensive patients randomized to pravastatin vs usual care. *JAMA* 2002;288:2998–3007.


100 Ballantyne CM, Houri J, Notarbartolo A, et al.: Effect of ezetimibe coadministered with atorvastatin in 628 patients with primary


CHAPTER 18

Metabolomics: applications in type 2 diabetes mellitus and insulin resistance

Sean H. Adams1,∗ and John W. Newman2
1Arkansas Children’s Nutrition Center, and Section of Developmental Nutrition, Department of Pediatrics, University of Arkansas for Medical Sciences, Little Rock, AR, USA
2Obesity & Metabolism Research Unit, United States Department of Agriculture-Agricultural Research Service, Western Human Nutrition Research Center; Department of Nutrition, University of California, Davis, CA, USA
∗Formerly affiliated with #2

Key points
• Type 2 diabetes mellitus (T2DM) involves perturbations in fuel metabolism beyond blood sugar. A classic example is the elevation in ketone body production that can progress to ketosis, indicative of abnormal mobilization and increased catabolism of free fatty acids. Modern approaches to detect, catalog, and interpret comprehensive suites of metabolites in blood or other biofluids (metabolomics) have highlighted that metabolic dysregulation also extends to amino acids and numerous other metabolic pathways.
• It is estimated that the human blood metabolome is composed of >6400 metabolites, and many routinely-detected molecules await final chemical annotation. A number of these non-annotated metabolites are altered by T2DM and track blood sugar control indices, suggesting that future advances in analytical chemistry and chemo-informatics will identify new markers of diabetes, prediabetes, and insulin resistance pathophenotypes and disease severity.
• Blood and tissue metabolite profiling supports the idea that in the T2DM and the insulin resistant states, there is a mismatch between fatty acid fuel delivery and combustion, for example in skeletal muscle. This leads to accumulation of lipid metabolites (diacylglycerol, ceramides, acylcarnitines) that may contribute to “lipotoxicity,” hypothesized to promote cell stress, inflammation, and insulin resistance.
• In addition to identifying biochemical pathways altered by T2DM, metabolomics analysis of blood and urine has helped identify predictive metabolite risk markers for T2DM. These markers may prove useful in establishing clinically viable diagnostic panels to predict disease stage or to help direct preventive strategies to thwart disease onset.
• Gut bacteria are implicated in regulating host metabolic health and themselves can be altered by host physiologic status and diet. Thus, considerations related to the human blood metabolome and its associations with metabolic disease should take into account the fact that some metabolites are derived from the gut microbiota. Additional research is needed to understand which, if any, of these “xeno-metabolites” are relevant to T2DM pathophysiology.

Introduction

Type 2 diabetes mellitus (T2DM) is not simply a disorder of blood sugar control, but a condition in which multiple biochemical pathways are impacted. An illustrative example of this is the elevated blood ketone body concentration typical of T2DM, which can progress to diabetic ketoacidosis if insulin action and/or availability deteriorates. The elevations of ketone bodies and the acetone accumulation (with associated “sweet breath odor”) of severe or poorly controlled diabetes are indicative of abnormal fatty acid homeostasis and specifically, high lipolysis and hepatic β-oxidation. These observations form the basis for the perspective that a primary event on the path to T2DM development is an aberration in lipid metabolism, that is, chronic exposure of peripheral tissues, liver and pancreatic β cells to high free fatty acids. According to this view, excess free fatty acids promote tissue insulin resistance, increase hepatic glucose output and act in pancreatic β cells to trigger hyperinsulinemia in response to insulin secretagogues [1,2]. While links between altered fatty acids and T2DM-relevant metabolic phenotypes have become well accepted, the mechanisms underlying the associations remain controversial, and a great deal is yet to be learned about other pathways and fuels that contribute to or are impacted by insulin resistance and diabetes. Broad-based metabolite profiling of blood and urine using targeted and untargeted analytical platforms can be a valuable approach in this regard.

By assessing how hundreds of individual metabolites are altered in human and animal models of T2DM, prediabetes, or insulin resistance, it has become evident that these conditions are often associated with alterations in fasting blood or tissue amino acids and their derivatives, systemic bile acids, blood phospholipids, and many as-yet non-annotated “unknown” metabolites. More research is needed to determine which of
these metabolite factors drive disease development and progression, and which are changed in response to deterioration of metabolic homeostasis. This is important not just from the standpoint of understanding basic pathophysiology and symptom etiologies in T2DM, but also to consider if select metabolites or combinations of metabolites have prognostic utility in terms of diabetes risk or disease staging. For instance, despite the clinical utility and cost-effectiveness of blood glucose measurement to diagnose T2DM (and more recently to define prediabetes), elevated blood glucose develops quite late in disease development and progression. Metabolomics tools can move the T2DM field beyond “glucocentric” medicine, to identify earlier biomarkers of disease risk, to foster therapeutic strategies targeting novel metabolic pathways, and ultimately to thwart diabetes or limit its progression and severity. For more comprehensive analysis of “omics” applications to insulin resistance and diabetes, the interested reader is referred to several recent reviews [3–5].

What is metabolomics?”

Metabolomics is an emerging field of analytical chemistry whose goal is to quantify all metabolites in a biologic sample, typically through application of multiple analytical platforms that enable determinations across a variety of metabolite classes (Figure 18.1). As with other omics-based technologies, this comprehensive coverage of metabolism has the benefit of allowing the observation of metabolic stressor effects not only at a single target, but also the ripple effect of the stressor, including compensatory responses, across the metabolic landscape. Metabolomics integrates and aggregates the information stemming from genomic, transcriptomic and proteomic events, providing access to a quantitative description of the metabolic phenotype of an individual (Figure 18.2). Moreover, while the genetic code and protein machineries responsible for metabolism provide the distinguishing characteristics between genera, species, and even individuals, the metabolites modified by this elegant and elaborate machinery are far less variable across disparate organisms. For example, all mammals use the same basic suite of metabolites to store, transport and generate energy, build membranes, and use in purine or pyrimidine turnover. Therefore, metabolomics provides a unique and powerful hypothesis-testing and hypothesis-generating tool in both the pre-clinical and clinical research environments.

The ability to discriminate individuals based on pattern of metabolites measured in biofluids has a long history, which has advanced along with technology [6]. An early biomedical application of this concept is newborn screening for inborn errors of metabolism. These tests employ diagnostic profiling of metabolites in blood spots (e.g. acylcarnitines, amino acids, organic acids) that can reveal tissue accumulation of intermediates resulting from specific metabolic lesions [7,8]. Metabolomics developed as an expansion of metabolic profiling, using pattern
The metabolome reflects the aggregate outcome of gene-to-whole body events. Determination of the comprehensive metabolite profile can provide insights into upstream enzyme activities and metabolic processes, each of which can be influenced by gene/protein expression levels and sequence polymorphisms. The metabolome is also influenced by the cellular environmental/biochemical milieu, metabolite degradation and excretion (not shown).

recognition to segregate subjects using both identified and unidentified components. However, the field is experiencing a transformation to a quantitative science as analyte peak identification and quantitative efficiency increases. As these efforts continue, repositories and searchable databases like those established within the Human Metabolome Database (www.hmdb.ca/) for metabolite reference ranges in the contexts of health and specific illnesses, will expand [10]. The blood and urine metabolome contains a broad array of metabolites, many of which are dynamically linked to concentrations in tissues and thus report on peripheral tissue metabolism. While less common, metabolomic studies of cerebrospinal fluid have also proven useful, offering novel insights into the progression of diseases such as amyotrophic lateral sclerosis [9]. Currently, the human metabolome is believed to contain at least 6400 unique chemical entities. While this number is impressive, many analytes observed in routine metabolomic screens remain unidentified (non-annotated). In the metabolomic paradigm, “unknowns” are considered useful and are assigned unique identifiers allowing their discrete identification in future studies. If such an unknown is later found to associate with a specific disease or treatment intervention, the effort to elucidate its structure may be warranted.

Mass spectrometry and proton nuclear magnetic resonance (NMR) spectroscopy are complementary techniques most routinely employed for metabolomic analyses, often using blood plasma or urine. While NMR is nondestructive, rapid, and inherently quantitative, it is also insensitive and as such best suited for higher-abundance compounds. On the other hand, while mass spectrometric methods are inherently more sensitive and able to discriminate complex mixtures by employing chromatographic separations, true quantification requires access to authentic standards. As with all information-rich technologies, the exploration, analysis and interpretation of metabolomics data is challenging. However, routine statistical approaches exist to effectively manage these data sets and distill them into discrete pieces of information. Metabolomics results are inherently data-heavy and complex, and to properly investigate them requires familiarity with multivariate statistical analyses approaches [11]. For instance, in case-control studies partial least squares discriminate analyses (PLS-DA) is often the default analysis approach. This process filters the data set to find those input variables that best characterize the differences between conditions (e.g. between case and control; pre- and post-treatment). While this approach to data reduction can remove 60–80% of the collected data to focus on variables that most robustly discriminate comparator groups, it can still leave hundreds of metabolites to consider. However, many of these compounds correlate strongly, and understanding which compounds move together (either increasing, or decreasing in concert) can further assemble the data into nodes of information that are descriptive of a biologic process. Hierarchical cluster analyses are the simplest approach to filtering data in this manner, and can parse the body of data into 5–10 pieces of discrete analyte groupings, a manageable amount to aid biologic interpretation. Another important issue that arises when dealing with such large data sets is the multiple comparison problem. Namely, using the statistical convention of a 95% confidence interval; if 100 measurements are made, 5 of these are expected to be statistically different by chance (i.e., \( \alpha = 0.05 \)). Various approaches have been employed to control for this “multiple comparison error.” The simplest and most conservative of these is the Bonferroni correction which limits the odds that a given observation is by chance by setting the \( p \)-value for significance at \( \alpha/n \). Therefore, if 1000 metabolites are measured and tested for significance, a \( p = 0.0005 \) would be required to claim a significant difference. A less conservative approach is controlling for the false discovery rate (e.g. [12]), which constrains the likelihood of making a false positive interpretation. However, an underlying assumption in these corrections is that the measured variables are independent and not collinear. In the case of metabolism, many metabolites are inherently linked. Thus, rather than a penalty, repeated determination of a difference in a group of related compounds should enhance interpretations as to the veracity of the results. In practice, this means that users of such data must carefully determine the true number of “independent” measurements that exist in a multivariate data set before application of these \( p \)-value adjustments.

Beyond the statistical analysis, visualization of these complex data sets is also an important tool for their analysis and interpretation. Where metabolic networks exist, fluxes and flows along metabolic pathways can be mapped. Similarly, correlation networks can be established showing interconnections between discrete pieces of information, thus helping to bridge and thus integrate the information into a more complete picture.
Altered metabolites and their potential origins in type 2 diabetes and insulin resistance

In recent years, a large number of cross-sectional and prospective studies have revealed specific metabolites and pathways that are altered by, or predictive of, T2DM and insulin resistance. Human insulin resistance, T2DM, and prediabetes involve significant perturbations in multiple metabolic systems, as reflected in positive associations with blood long- and medium-chain acylcarnitines, branched-chain amino acids (BCAAs) or their derivatives, methionine/threonine-derived 2-hydroxybutyrate (2-HB), phenylalanine or tyrosine, and negative associations with blood glycine and select phospholipids [13–27]. These metabolites reflect underlying processes that are impacted by or participate in T2DM or insulin resistance pathophysiology. Based on current knowledge, however, it is unclear which, if any, of the metabolites themselves contribute as a root cause of disease.

As noted above, a perturbation in fatty acid metabolism is a hallmark of insulin resistance and T2DM. The weight of evidence from metabolomics studies and tissue biochemical determinations points to a mismatch of fatty acid fuel delivery (due to elevated lipolysis and circulating fatty acids) relative to complete combustion in mitochondria in situ [28]. This leads to accumulation of muscle and blood long- and medium-chain fatty acylcarnitine metabolites that report on tissue fatty acyl-CoA status [14,28]. The tissue origins of higher systemic blood acylcarnitines in T2DM remain to be confirmed, but muscle is a known site of robust acylcarnitine generation. However, mitochondrial mismatch and the incomplete oxidative catabolism of fatty acids likely occurs in all tissues including liver, supported by consistent observations of higher blood ketone bodies in T2DM. Other metabolite derivatives that result from excess tissue fatty acid availability and that have been proposed to be associated with insulin resistance phenotypes include ceramides and diacylglycerols [29].

Both targeted metabolic profiling and global metabolomics experiments have indicated that amino acids should be considered alongside glucose and lipids in the context of the dysmetabolic insulin resistant or T2DM states. In the late 1960s it was observed that BCAAs are increased in the blood of obese, insulin-resistant subjects [30], but only recently did comprehensive metabolite profiling place BCAAs into the spotlight as metabolites associated with insulin resistance [31]. Other essential amino acids such as phenylalanine, tyrosine, methionine and/or the methionine derivatives cysteine-cystine and 2-HB have also been reported to be modestly increased in blood from T2DM and/or insulin resistance conditions (reviewed in [32]). In the case of associations of elevated blood amino acids with insulin resistance and T2DM phenotypes, a growing body of evidence suggests that reductions in tissue utilization and/or incomplete oxidation (due to lower enzyme activities) could play a role. For instance, insulin resistance in a human cohort (highest insulin concentration) was associated with significantly attenuated post-oral glucose tolerance test (OGTT) reductions in blood leucine/isoleucine [33], supporting the idea that lower insulin-stimulated tissue BCAA utilization occurs with insulin resistance. Lower expression of mitochondrial branched chain ketoacid dehydrogenase complex (BCKDC, main control point of BCAA oxidative catabolism) and associated proteins and genes is evident in visceral white adipose tissue from obese, insulin-resistant rodents and humans [34,35] (and references therein). Reduced expression of BCKDC, and lower enzyme activity due to the high fatty acid environment of T2DM and insulin resistance, could promote increased concentration of upstream metabolites such as BCAAs, methionine/cysteine, and 2-HB/2-ketobutyrate [32]. Recently, metabolomics analysis of blood from a large prospective study cohort (Framingham Heart Study) revealed the lysine derivative 2-aminoadipic acid (2-AAA) as predictive of T2DM development [36], with persons in the highest quartile of 2-AAA concentrations at fourfold greater risk. Lower blood glycine concentrations have also been reported to associate with insulin resistance or T2DM risk (e.g. [19,20,24,25]). The origins of the latter metabolic phenotypes remain unknown.

Additional metabolite classes that are coming to the fore in terms of their potential association with insulin resistance or T2DM risk include bile acids and choline-containing phospholipids. For instance, insulin-resistant individuals had a blunted increase in blood glycochenodeoxycholic acid following OGTT [33]; yet, fasting concentration of this metabolite was increased in a cohort of impaired glucose tolerant vs. normal glucose tolerant individuals [18]. Phosphatidylycholine derivatives including glycerophosphatidylcholine metabolites are consistently reported to be reduced in T2DM or insulin resistance [16,18,20,25,37]. Further research is needed to determine why insulin resistance and T2DM phenotypes associate with blood patterns of bile acids and phospholipids, but it is possible that changes to the gut environment and/or gut bacteria (microbiota) play a role. Recent metabolomics results highlight that the host’s metabolic health and insulin resistance status are associated with differences in profiles of blood and urine xenometabolites (“nonself” metabolites) (e.g. [18,38]). Interestingly, taxonomic diversity in the gut microbiome is significantly different in persons displaying risk markers for cardiometabolic disease including insulin resistance [39], and there are clear differences in gut microbe communities when comparing subjects with normal glucose tolerance, prediabetes, and T2DM [40–43].

Future directions and potential application of metabolite profiling in type 2 diabetes

Metabolomics approaches have unmasked a variety of metabolic pathways influenced by T2DM or insulin resistance, moving the field of metabolic physiology beyond a “glucocentric” viewpoint of these conditions. However, a difficulty in interpretation is
that the fasting concentration of a metabolite only reflects a single moment in time, and etiology or tissue origins of any metabolite differences across comparator groups requires further validation. To fully elaborate the molecular and physiologic mechanisms underlying metabolite differences, future efforts will conduct systematic exploration into relevant enzymology, biochemistry, and flux measures (i.e., metabolite turnover, anabolism-catabolism events). Furthermore, efforts are underway to determine which systemic metabolites associated with metabolic diseases such as T2DM emanate from gut microbes or co-metabolism of microbe and host.

From the standpoint of clinical practice, what does metabolomics have to offer? In the future, determination of metabolic variants in drug metabolism in the clinic could be used to tailor therapeutic regimens, revolutionizing the application of “pharmacometabolomics” to individualized medicine [44]. If applied as a tool to routinely characterize patient health, metabolomics should provide novel diagnostic and prognostic tools allowing early disease identification and intervention. While fasting samples should prove valuable in this regard, it is likely that dynamic changes in one or more analytes following OGGT or other challenges will prove even more sensitive. Comprehensive metabolite profiling of blood and tissues has already led to at least one first-generation analyte panel relevant to T2DM and metabolic disease research and medical evaluation. For instance, Quantose™ is a fasting blood analyte panel available for research purposes from Metabolon (Research Triangle Park, North Carolina, USA), and includes the analytes 2-HB, 1-lineoyl-glycerophosphocholine (L-GPC), oleic acid, and insulin. The panel score reflects relative risk for T2DM and may be useful for experiments involving relative stage or severity of disease progression, or those examining efficacy of interventions. The same panel is also incorporated into a Diabetes Prevention and Management Panel offered by Health Diagnostic Laboratory (Richmond, Virginia, USA). These early examples of tests that are based on multi-analyte platforms have their foundation in pioneering cross-sectional and prospective cohort metabolomics research. With advancements in technology and clinical uptake, it is feasible that robust predictive tests for T2DM will become commonplace, helping guide prevention strategies well before disease onset, and treatment strategies to minimize disease severity.

References


SECTION IV

NAFLD, NASH and non-traditional associations with diabetes
CHAPTER 19
Pathogenesis of nonalcoholic fatty liver disease (NAFLD)

Hannele Yki-Järvinen
Department of Medicine, University of Helsinki, Minerva Foundation Institute for Medical Research, Helsinki, Finland

Key points
• Nonalcoholic fatty liver disease (NAFLD) is the most common liver disorder.
• NAFLD covers a range from simple steatosis (NAFL) to nonalcoholic steatohepatitis (NASH).
• NASH may progress to cirrhosis and end-stage liver disease.
• NAFLD is the most common cause of hepatocellular carcinoma (HCC).
• NAFLD predicts type 2 diabetes and cardiovascular disease independent of obesity.
• Obesity and increased simple sugar intake are the major acquired risk factors for NAFLD.
• The fatty liver due to NAFLD is insulin resistant and overproduces glucose and VLDL, which results in hyperinsulinemia and low HDL cholesterol.
• Carriers of the I148M variant in PNPLA3 (PNPLA3 NAFLD) have an increased risk of NASH and HCC but do not have features of the “metabolic syndrome” (MetS).

Introduction
NAFLD is defined as steatosis (>5–10% of hepatocytes are fatty), which is not due to excess use of alcohol (defined in European and American guidelines as >20 g of alcohol daily for women and >30 g for men), or other conditions as determined by careful family and medical history, and laboratory tests to exclude at least steatosis due to viral and autoimmune causes and iron overload [1].

NAFLD is usually asymptomatic and most patients have normal transaminases (ALT <30–40 U L⁻¹ for men and <20–30 U L⁻¹ for women) [2,3], although NAFLD is the commonest cause of incidentally discovered elevated liver function tests [4] and the most common cause of chronic liver disease [5] (Figure 19.1). The prevalence of NAFLD depends on the method of diagnosis but is as common as the MetS (metabolic syndrome). The estimated worldwide prevalence is in the range from 15% (China) to 30% (Dallas Heart Study) [6].

Some patients with NAFLD have NASH, which is characterized in addition to steatosis by ballooning necrosis in the vicinity of steatotic hepatocytes, mild inflammation, and possibly fibrosis [7]. NASH can only be diagnosed by a liver biopsy. Recent reviews have estimated the worldwide prevalence of NASH to be 3–6% [6,8]. Although the main cause of excess mortality in NASH is CVD [8], NASH may progress to cirrhosis and end-stage liver disease. Cirrhosis may occur in ~5% of patients with NASH (see [8] for review) (Figure 19.2). NASH is perhaps the most common cause of cryptogenic cirrhosis [9]. It is currently the third most important indication for liver transplantation in the US and the only cause with an increasing incidence [10]. Recent data have shown that NAFLD is the most prevalent cause of HCC [11] (Figure 19.1), which may occur even without cirrhosis in patients with NAFLD [12,13]. The risk of HCC is increased at least twofold in obesity and type 2 diabetes [13] and may be especially high in patients with NAFLD, who are carriers of the common variant in PNPLA3 (see later) [14]. Identification of patients at risk for NASH and HCC is a major challenge.

The liver is the site of production of two of the key components of the MetS, fasting serum glucose and very-low density lipoprotein (VLDL), which contains most of the triglycerides present in serum. In subjects with NAFLD, the ability of insulin to normally suppress production of glucose and VLDL is impaired resulting in mild hyperglycemia and hypertriglyceridemia, two key features of the MetS [15,16]. The liver, once fatty, also overproduces other markers of cardiovascular risk such as C-reactive protein (CRP), fibrinogen, and coagulation factors [8,17]. The ensuing chapter is focused on reviewing the pathogenesis of NAFLD mostly based on human studies.
Pathogenesis of NAFLD

Sources of intrahepatocellular triglycerides
The hallmark of NAFLD is steatosis, that is, accumulation of triglycerides within hepatocytes. The fatty acids in these triglycerides can be derived from peripheral lipolysis, that is, fatty acids released from adipose tissue, de novo synthesis of fatty acids from simple sugars such as glucose and fructose as well as from amino acids via the pathway of de novo lipogenesis (DNL), or from dietary fatty acids, which can reach the liver via the spillover pathway and through uptake of intestinally derived chylomicron remnants [18] (Figure 19.3). Spillover is defined as the amount of free fatty acids (FFA) that are released from triglycerides during intravascular lipolysis but not taken up by adipose or muscle tissues. Studies using stable isotopes in humans have suggested that after an overnight fast and during the night, lipolysis provides approximately half of the fatty acids stored in intrahepatocellular triglycerides [19]. The contribution of lipolysis to intrahepatocellular triglycerides does not appear to be increased if weight-matched subjects with and without
Pathogenesis of nonalcoholic fatty liver disease (NAFLD)

Figure 19.2 NAFLD as a predictor of disease. Once fatty due to nonalcoholic causes, often in subjects with the metabolic syndrome (MetS), the liver is insulin resistant and overproduces many factors such as glucose, very-low-density lipoprotein (VLDL), C-reactive protein (CRP), plasminogen activator inhibitor 1 (PAI-1), fibrinogen, coagulation factors (FVIII, XI, XII) and cholesterol. In some subjects, perhaps 5%, nonalcoholic steatohepatitis (NASH) develops. NASH predisposes to cirrhosis and hepatocellular carcinoma. Obesity, type 2 diabetes and insulin resistance as well as the I148M allele in PNPLA3 are predictors of NASH. NAFLD also increases, independent of obesity, the risk of cardiovascular disease, deep venous thromboembolism, and cholesterol gallstones.

Figure 19.3 Sources of intrahepatocellular triglycerides (TAGs) in NAFLD. Dietary fat is transported from the gut in chylomicrons in lymph via the thoracic duct to the circulation. Chylomicrons undergo intravascular lipolysis (1) during which part of the FFA are released by lipoprotein lipase (LPL) at the capillary endothelium and are taken up by adipose and muscle tissues. During this process, some of the FFA spill over to the systemic circulation where they are available for uptake by tissues such as the liver (3). After intravascular lipolysis, chylomicron remnants are taken up by the liver (4). Fatty acids (FA) can also be synthesized from dietary sugars and amino acids via the de novo lipogenesis (DNL) pathway (2). In the liver, part of the TAGs are routed for VLDL assembly. Lipolysis of TAG to FA may occur prior to TAG synthesis for VLDL. Source: Donnelly 2005 [18]. Reproduced with permission of American Society for Clinical Investigation.
increased liver fat content due to NAFLD are compared [19]. Despite this, multiple studies have shown adipose tissue lipolysis to be increased in subjects with NAFLD [20,21]. The pathway, which seems most abnormal in NAFLD is DNL [19]. During a meal, the contribution of lipolysis decreases while that of dietary fatty acids increases [18]. Another mechanism by which triglycerides can accumulate is defective lipolysis within hepatocytes. A major form of NAFLD due to genetic variation in PNPLA3 has recently been identified. The 1148M variant in PNPLA3 may impede breakdown of intrahepatocellular triglycerides. This cause of NAFLD as well as others and the mechanisms by which they contribute to NAFLD are discussed later.

**Cause of NAFLD: obesity and abnormalities in adipose tissue**

Although both NAFLD and the MetS can occur in nonobese subjects, the prevalence of NAFLD is markedly increased in obesity as is that of the MetS. In the third National Health and Nutrition Examination Survey (NHANES), the prevalence of NAFLD averaged 7.5% and 6.7% in normal-weight men and women but was 57% and 44% in persons with a body mass index (BMI) > 35 kg m\(^{-2}\) [22]. Asians have a higher prevalence of NAFLD than Caucasians for a given BMI. For example, the mean BMI in 6905 Chinese adults with ultrasound-diagnosed NAFLD was 23.6 kg m\(^{-2}\) as compared to 21.5 kg m\(^{-2}\) in those without NAFLD [23]. Obesity is a true cause of NAFLD and hepatic insulin resistance as weight loss rapidly reverses both [see [24] for review]. Bariatric surgery induced remission regrseses histologic changes, at least steatosis and ballooning, in NASH [25].

In obese subjects, who have NAFLD, adipose tissue is characterized by multiple alterations such as hypoxia (see [26] for review), increased infiltration of macrophages surrounding dead adipocytes and expression of chemokines and pro-inflammatory cytokines [27], fibrosis as well as deficiency of adiponectin [28]. Inflamed adipose tissue is insulin resistant leading to increased FFA flux to the liver especially under conditions mimicking the postprandial state [29]. In obese subjects, a decrease in serum adiponectin is likely to contribute to fat accumulation in the liver and may also induce hepatic inflammation and insulin resistance (see [28] for review). Changes in gut microbiota have been suggested to contribute to the pathogenesis of both obesity and NAFLD syndrome (see [30] for review). The change in intestinal bacterial flora is thought to alter the permeability of gut mucosa to inducers of inflammation such as endotoxin. In addition to abnormal adipose tissue metabolism, an increase in simple sugar intake and DNL characterizes obese a compared to nonobese subjects (see later).

Over 50 years ago, Jean Vague classified obese subjects according to the degree of “masculine differentiation” [31] into those with “gynoid” and those with “android” obesity. Gynoid obesity was characterized by lower-body deposition of fat (around the thighs and buttocks, “pear-shaped”) and relative underdevelopment of the musculature, while android obesity defined upper-body (truncal, “apple-type”) adiposity, greater overall muscular development and a tendency to develop hypertension, diabetes, atherosclerosis, and gout. These phenotypic observations have subsequently been confirmed in many cross-sectional and prospective studies (see [32] for review). The amount of visceral fat correlates closely with the amount of fat in the liver [3].

Although abdominal obesity is a marker of NAFLD and the MetS, it is controversial whether visceral fat is “causal or correlative” [33] or “the major culprit or an innocent bystander” [34] for features of the MetS or NAFLD. The small visceral depot has been suggested to be harmful because omental and mesenteric adipocytes have a higher rate of lipolysis than subcutaneous ones, and this could liberate excessive amounts of FFA into the portal vein and make the liver fatty. Visceral fat may also release more inflammatory cytokines than subcutaneous adipose tissue [32].

However, this “portal theory” can be questioned [33]. When groups of subjects, who have similar amounts of liver fat but differ with respect to the amount of subcutaneous and visceral fat, are compared, there are no defects in insulin suppression of lipolysis or in rates of VLDL production [35]. If overactive lipolysis were responsible for the harmful effects of visceral fat, one would predict this depot to be reduced rather than enlarged. Surgical removal of omental fat in humans does not improve hepatic insulin sensitivity [36].

Patients with NAFLD frequently have obstructive sleep apnea (OSA). According to two meta-analyses pooling data from 18 [37] and 11 [38] studies, patients with OSA have a significantly and 2.0–2.6-fold increased risk of having NAFLD and a significantly higher risk of liver fibrosis independent of obesity [39]. This increased risk could be mediated by hypoxia (see [26] for review). Patients with the polycystic ovary syndrome are more insulin resistant and have a higher liver fat content than equally obese subjects without the syndrome [40].

**Cause of NAFLD: high simple sugar intake and de novo lipogenesis**

The increase in the consumption of foods containing high amounts of added simple sugars such as soft-drinks, deserts, cookies, sugar and candy, and fruit drinks has far exceeded that in intake of any other food group [41]. Recent prospective epidemiologic data have shown that an increase in the intake of sugar-sweetened beverages (SSB) predicts both obesity, type 2 diabetes, CVD, and NASH independent of other factors such as energy intake. In the HPFU (Health Professionals Follow-Up Study), the risk of CVD was significantly increased in men using high amounts of SSB after adjusting for confounders such as physical activity and energy intake. Intake of SSB was also significantly associated with features of the MetS such as hypertriglyceridemia, low HDL cholesterol, and high CRP and IL-6 concentrations [41]. There were similar findings in the Nurses’ Health Study. Analysis of food questionnaires from 427 adults with biopsy-proven NASH found fructose consumption, after adjusting for age, gender, BMI and total caloric intake, to be significantly associated with fibrosis [42]. Taken together
these data suggest that simple sugars significantly contribute to the epidemics of obesity, type 2 diabetes, CVD, and NASH.

Simple sugars are converted to fat in the liver via DNL [43]. This pathway accounts for some 5% of liver fat synthesis in normal subjects but is threefold higher in patients with NAFLD when compared to equally obese subjects without an increased liver fat content [19]. It produces exclusively saturated fatty acids [43], which relative proportion in the liver [44], as well as rate of production increases in direct proportion to liver fat in humans [45]. Saturated fats produced by DNL as compared to fatty acids derived from the diet may cause more lipotoxic injury (see [46] for review). Some of fructose may be metabolized in the colon and influence gut microbiome [46]. Overfeeding a high carbohydrate (1000 extra simple sugar calories as candies and sugar-sweetened beverages/day) diet for merely 3 weeks increases DNL, liver fat (by 30%), and serum triglycerides and lowers HDL cholesterol [47]. Meta-analyses comparing isocaloric substitution of carbohydrate by fat (saturated or monounsaturated or polysaturated) have shown that high carbohydrate but not high fat diets produce the dyslipidemia typical of the MetS and NAFLD [48]. The rate of DNL is increased in obese subjects who are hyperinsulinemic but not in those who are normoinsulinemic, even when obese subjects are consuming the same diet as nonobese subjects [49]. These data imply that excessive intake of simple sugars recapitulates key features of the MetS as well as NAFLD, and that DNL is also increased in hyperinsulinemic subjects irrespective of diet composition.

Cause of NAFLD: physical inactivity
Although physical training predominantly enhances muscle insulin sensitivity, cross-sectional epidemiologic data and studies on effects of physical training suggest that exercise may decrease liver fat even in the face of unchanged body weight [50]. In a cross-sectional study in 72,359 healthy Korean adults, subjects who were exercising regularly had a 28–53% lower risk of NAFLD across almost all BMI deciles [51]. In a small (n = 12) but carefully performed mechanistic study, exercising 30–60 min five times a week for 16 weeks modestly decreased liver fat by 10% in the absence of changes in body weight [52]. The effect of exercise seems small in comparison to the effects of a hypocaloric ketogenic diet; when the same group followed a low carbohydrate diet, liver fat was shown to decrease by 30% within 48 hours [53]. Exercise may reduce liver fat by channeling dietary carbohydrates to muscle glycogen away from DNL [54].

Causes of NAFLD: familial and genetic factors

Family history, ethnicity
NAFLD has been reported to run in families with heritability estimates of around 40% [55]. This familial clustering is observed even after adjusting for age, gender, and obesity [55]. The heritability of S-ALT due to nonalcoholic causes based on a study in twins was 55% [56]. Ethnicity influences the prevalence of NAFLD [57]. In the Dallas Heart Study, 45% of Hispanics, 33% of Whites, and 24% of Blacks had NAFLD [2]. These differences were only in part attributable to differences in BMI. As discussed later, genetic variation in PNPLA3 may contribute to ethnic variation in the prevalence of NAFLD.

Age and gender
In the third NHANES, the prevalence of NAFLD increased linearly from subjects less than 30 years to those aged 50–60 years and was higher in men than in women.

The PNPLA3 I148M variant
A common genetic variant, an allele in PNPLA3 (rs738409[G], encoding I148M) has been found to be a major factor in determining liver fat content and susceptibility to NASH, independent of obesity. This variant was discovered in 2008 in a genome-wide association scan in Hispanic, African American, and European American individuals. It showed genetic variation in PNPLA3 to confer susceptibility to NAFLD. An allele in PNPLA3 (rs738409[G], encoding I148M) was associated with increased liver fat and also hepatic inflammation and fibrosis [58]. This finding has subsequently been reproduced. In a meta-analysis comprised of 16 studies published in 2011, homozygous carriers of the variant as compared to noncarriers had a 73% higher lipid fat content, a 3.2-fold greater risk of high necro-inflammatory scores and a 3.2-fold greater risk of developing fibrosis [59] (Figure 19.2). The association between the PNPLA3 variant allele and steatosis has been observed in seven out of eight GWAS studies and in multiple other studies including Chinese, African Americans, participants of the third NHANES, children with NAFLD, and morbidly obese patients [60]. The association between the PNPLA3 rs738409[G] allele and severity of histologic liver disease has also been confirmed [60]. As summarized in a meta-analysis, recent studies have also shown that the PNPLA3 rs738409[G] allele is associated with an increased risk of HCC [14]. In the population-based Dallas Heart Study, the prevalence of the PNPLA3 rs738409[G] allele displayed ethnic variation, being highest in Hispanics and lowest in African Americans [58]. The PNPLA3 I148M variant is indeed common. In subjects with NAFLD diagnosed by US or 1H-MRS, the prevalence of subjects homozygous vs. heterozygous for the rs738409[G] allele has averaged 12 vs. 39% in Europeans [61], 17 vs. 50% in Chinese [62], and 19 vs. 51% in Taiwanese children [63]. In subjects without NAFLD, the prevalences were 8 vs. 31%, 11 vs. 47%, and 12 vs. 48%, respectively.

In humans, PNPLA3 is expressed predominantly in the liver. Initial in vitro studies using purified human PNPLA3 showed that the wild-type enzyme hydrolyzes triglycerides and that the I148M substitution abolishes this activity [64]. These data suggest that the I148M substitution is a loss-of-function mutation impairing triglyceride hydrolysis. However, PNPLA3 also stimulates lipogenesis in vitro and the PNPLA3 I148M variant was suggested to increase this activity [65]. Thus, there are still
uncertainties as to how the gene variant influences the biology of human NAFLD.

In almost all studies, the NAFLD associated with PNPLA3^{148MM/148MI} was not accompanied by either insulin resistance, hyperglycemia, hypertriglyceridemia, or a low HDL cholesterol concentration, or with inflammation in adipose tissue [66]. On the other hand, it is important to recognize that the same person may have NAFLD due to obesity/insulin resistance and NAFLD due to the PNPLA3 I148M gene variant. Genotyping for PNPLA3 at rs738049 may become part of screening strategies in the clinic for patients with steatosis as this may help in predicting risk of NASH and HCC.

**Consequences of NAFLD**

**Hyperglycemia and hyperinsulinemia**

After an overnight fast, the main action of insulin is to restrain endogenous hepatic glucose production, while the rate of glucose uptake is largely insulin-independent [67]. In subjects with NAFLD [15,21] the ability of insulin to inhibit glucose production is impaired resulting in mild hyperglycemia. The latter stimulates insulin secretion leading to hyperinsulinemia. Fasting hyperglycemia and hyperinsulinemia are thus consequences of hepatic insulin resistance and correlate positively with liver fat, even independent of BMI (Figure 19.4).

Nondiabetic hyperglycemia and hyperinsulinemia are well established predictors of type 2 diabetes. Consistent with such data and the close association between liver fat and these predictors, NAFLD diagnosed by US has been shown to predict type 2 diabetes in multiple prospective studies (see [8] for review). The association between NAFLD and type 2 diabetes was independent of obesity in all studies, which adjusted for obesity. NAFLD predisposes to type 2 diabetes by increasing the amount of insulin needed to keep hepatic glucose production in the nondiabetic range. However, in subjects predisposed to develop type 2 diabetes, β cells are unable to sustain insulin secretion leading to relative β cells failure and overt type 2 diabetes.

![Figure 19.4](image-url) Consequences of hepatic insulin resistance. Upper panel: Insulin normally restraints hepatic glucose production. Once fatty, the ability of insulin to inhibit glucose production is impaired, which stimulates insulin secretion leading to hyperinsulinemia. Lower panel: Insulin normally restraints hepatic VLDL production. Once fatty, the ability of insulin to inhibit glucose production is impaired, which stimulates insulin secretion leading to hyperinsulinemia.
Some studies have suggested that NAFLD predicts type 2 diabetes even independent of all components of the MetS and thus that NAFLD is an even better marker of risk than the MetS.

**Hypertriglyceridemia and low HDL cholesterol, increased risk of type 2 diabetes**

Under fasting conditions, the liver of subjects with NAFLD overproduces triglyceride-enriched VLDL particles despite hyperinsulinemia when compared to equally obese subjects without NAFLD [68]. Insulin normally decreases production of VLDL by inhibiting adipose tissue lipolysis, and by directly suppressing hepatic production of VLDL [69]. In subjects with NAFLD, insulin fails to suppress both lipolysis [20] and production of triglyceride-rich VLDL particles from the liver [16]. This hepatic overproduction of large triglyceride-rich VLDL particles is another consequence of hepatic insulin resistance and the major contributory mechanism underlying the increase in serum triglycerides in subjects with the Met and NAFLD [69] (Figure 19.4). The increase in VLDL leads to lowering of HDL cholesterol and also to generation of small dense LDL particles, which are known to be highly atherogenic (see [32] for review).

**Markers of cardiovascular risk, increased risk of cardiovascular disease and thromboembolic complications**

The liver in subjects with NAFLD overproduces multiple markers of cardiovascular risk. These include, in addition to liver enzymes, glucose, and VLDL, increased production of C-reactive protein (CRP), fibrinogen, coagulation factors (FVII-IX, XI-XII), and plasminogen activator inhibitor-1 (PAI-1), and decreased production of insulin-like growth factor binding protein 1 (IGFBP-1) and sex-hormone binding globulin (SHBG) [70]. Changes in these circulating markers could explain, at least partly why NAFLD predicts CVD (see [8] for review), also when adjusted for obesity [8]. The changes in the markers of fibrinolysis and coagulation could contribute to an increased risk of thromboembolic complications in NAFLD [17].

**Other metabolic disease**

NAFLD increases the risk of cholesterol gallstones [41]. The fatty liver overproduces cholesterol [71] and displays other alterations in cholesterol metabolism [72], which may contribute to formation of cholesterol gallstones and also hepatic damage (Figure 19.2). NAFLD is also associated with an increased prevalence of obstructive sleep apnea and polycystic ovary syndrome.

**Conclusions**

NAFLD is a new epidemic, which has important health consequences. The data summarized in this review support the idea that NAFLD, which is not due to genetic variation in PNPLA3 “metabolic NAFLD,” closely resembles the MetS with respect to its causes and consequences. For these subjects, liver fat content is an excellent marker of the metabolic abnormalities characterizing the MetS. Indeed, since NAFLD predicts type 2 diabetes, even independent of the MetS, it may be better than the MetS in predicting risk. In the absence of simple and reliable tools to diagnose NAFLD, screening for simple steatosis is not recommended at present [1]. However, an incidental finding of NAFLD should lead to careful evaluation of the presence of risk factors for diabetes and CVD as well as for NASH. NAFLD may also be encountered more often than by chance in patients with otherwise unexplained deep venous thromboembolism and gallstone disease. On the other hand, the possibility of NASH should be remembered in all patients with the MetS by calculating appropriate risk scores such as the NAFLD liver fibrosis score [73,74]. PNPLA3 genotyping may become a routine test in all subjects with NAFLD as these patients are at particularly high risk of NASH and even HCC but not diabetes and CVD. However, the clinical utility of such testing still warrants further evaluation.

**References**


19 Lambert JE, Ramos-Roman MA, Browning JD, Parks EJ: Increased de novo lipogenesis is a distinct characteristic of individuals with nonalcoholic fatty liver disease. *Gastroenterology* 2014;146(3):726–735.


CHAPTER 20

Treatment of nonalcoholic fatty liver disease (NAFLD) and nonalcoholic steatohepatitis (NASH)

Fernando Bril, Kwame Ntim, Romina Lomonaco and Kenneth Cusi
University of Florida, Gainesville, FL, USA

Key points
• Nonalcoholic fatty liver disease (NAFLD), and its more severe form with hepatocyte necro-inflammation and eventual lobular fibrosis called nonalcoholic steatohepatitis (NASH), are much more common in clinical practice than previously thought. This is in part due to the increase in obesity and T2DM.
• Lifestyle intervention remains the gold-standard treatment for NAFLD and NASH.
• Although the best lifestyle strategy to address NAFLD/NASH is not currently known, any intervention leading to significant (>5–10%) weight reduction (i.e., diet, exercise, bariatric surgery) decreases liver triglyceride accumulation, improves liver histology and ameliorates the associated comorbidities (dyslipidemia, insulin resistance, hyperglycemia).
• Of the many drugs studied in clinical trials for the treatment of NASH, only vitamin E, obeticholic acid and pioglitazone offer clear evidence of benefit from RCTs in terms of safety and efficacy in adults with NASH.
• New pharmacologic agents are undergoing clinical trial testing and offer promise for the management of patients with NAFLD/NASH.

Introduction

Nonalcoholic fatty liver disease (NAFLD) encompasses a spectrum of liver disease ranging from simple liver fat accumulation to severe hepatic necro-inflammation (nonalcoholic steatohepatitis or NASH), and eventually cirrhosis [1]. Nonalcoholic fatty liver disease has become the most common cause of chronic liver disease and its prevalence will continue to increase as a consequence of the obesity epidemic. According to a study by Szczepaniak et al. [2] using the current gold-standard magnetic resonance imaging and spectroscopy (MRS) technique, about one third of the adult US population and as many as two thirds of obese adults have NAFLD. More importantly, a significant number of patients are believed to have NASH. The real proportion remains uncertain depending on the population studied, clinical setting, and diagnostic screening approach. In patients screened for NAFLD in tertiary medical care settings, the prevalence of liver biopsy-proven NASH among patients with steatosis ranges from 30% to 44% [3,4].

Previously thought to be a harmless condition associated with the metabolic syndrome, NAFLD and NASH are now recognized as important risk factors for cardiovascular disease [5,6], type 2 diabetes mellitus (T2DM) [7], cirrhosis [8], and hepatocellular carcinoma [9]. This emphasizes the importance of early diagnosis. However, the diagnosis of NASH can be challenging as it depends on performing a liver biopsy, which many primary care physicians and patients are unwilling to undergo.

Once NASH is diagnosed in a patient with T2DM, the main issue is how this diagnosis will affect the patient’s overall management. At present, there is no drug specifically approved for the treatment of NAFLD or NASH. While some pharmacologic agents appear to be promising [10,11], lifestyle intervention (mainly diet plus exercise) remains the standard of care to avoid disease progression [12,13].

Several drugs have been assessed in clinical trials for the treatment of NASH [10,11,14–19]. However, most of the studies have been small, of short duration, or had relied on surrogate markers (liver enzymes or ultrasound) for the diagnosis of NAFLD and for outcome measures. Only vitamin E (in patients without T2DM) and pioglitazone (in both patients with or without T2DM) have proven to be safe and effective in randomized, placebo-controlled trials [10,11]. Ongoing studies are assessing novel drugs for NAFLD/NASH but clinical use may be a few years away at the present time. Therefore, this chapter focuses primarily on the available evidence of nonpharmacologic and pharmacologic approaches for NAFLD and NASH.
Role of lifestyle intervention in NAFLD

It has been shown that hypocaloric diets, with or without physical activity, reduce the risk of cardiovascular disease (CVD) and T2DM [20,21]. However, their effect in patients with NAFLD has been less well studied. Most studies have had serious shortcomings ranging from small sample size to short duration [22–28], or simply lacking a control group [29–31]. Even in strongly controlled studies, significant variability in the dietary or exercise protocols make comparisons between studies and generalized conclusions difficult [32,33]. Another problem is that in the majority of trials, the diagnosis of NAFLD and treatment effect were not assessed by means of liver histology (liver biopsy), but rather with the use of surrogate markers [34–38]. Beyond these limitations, there appears to be consensus that patients with NAFLD benefit from diet and physical activity [13], although it is still unclear which lifestyle intervention strategy is the most beneficial. Studies have been inconsistent when assessing different types of diet (low fat, low carbohydrates, and so on) [22–24,39,40], and exercise training protocols (i.e., aerobic vs. resistance training) [41], either with or without weight reduction. In the following sections we will discuss efforts in trying to establish the best approach for patients with NAFLD/NASH.

Dietary alone interventions

The effect of dietary intervention alone for the management of NAFLD and NASH has been evaluated extensively over the past 2–3 decades [42,43]. These studies can be divided into two main types: those that have focused on weight reduction only, and clinical trials that have assessed the role of a particular dietary composition on hepatic steatosis.

Several small and uncontrolled studies have examined the effect of diet-induced weight reduction (without physical activity changes) on plasma aminotransferases levels and hepatic steatosis measured by either liver ultrasound [34,44] or computed tomography (CT) [35,36]. Taken together, a 4–5% weight loss by dietary intervention alone usually improves plasma aminotransferases levels and hepatic steatosis. Unfortunately, these surrogate markers of liver disease have been shown to have weak correlations with histologic findings [45], and do not allow to make firm conclusions regarding the long-term value of weight reduction by dietary intervention alone in NAFLD.

Table 20.1 summarizes the studies that have assessed the effect of weight reduction by diet alone on hepatic triglyceride accumulation by MRS in patients with NAFLD. As can be observed, independently of the type of diet used, weight reductions ranging from 6.5 to 11.0% have been associated with a significant reduction in liver fat content by MRS in the order of 38% to 81%. Moreover, Tendler et al. [46] in a provocative small (n = 5) pilot study reported that after 6 months of a low-carbohydrate, ketogenic diet, a reduction of ~11% of body weight was associated with qualitative improvements in liver histology.

Only a few studies have examined the role of dietary composition in addition to negative caloric balance. Ryan et al. [47] reported significant decrease in plasma aminotransferase levels with a low-carbohydrate diet when compared to a high-carbohydrate regime in patients with T2DM. However, interpretation is difficult as this improvement was associated with a larger weight reduction in the low-carbohydrate diet group. Kirk et al. [39] found similar reductions in liver fat content and weight after 11 weeks of a low-carbohydrate (high-fat) vs. high-carbohydrate diet. Of interest, short-term assessment after 48 hours of calorie restriction already showed a significant reduction in LFAT (~20%) in the low-carbohydrate diet when compared to the high-carbohydrate arm, even before any meaningful weight change. In 170 overweight and obese patients with NAFLD treated over a 6-month period with a hypocaloric diet restricting either carbohydrates or fat, Haufe et al. [40] concluded that both interventions were equally

<table>
<thead>
<tr>
<th>Author [ref.] (year)</th>
<th>Type of study</th>
<th>Duration (weeks)</th>
<th>Type of hypocaloric diet</th>
<th>Weight loss</th>
<th>Change in liver fat by MRS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n = 8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kirk [39] (2009)</td>
<td>RCT</td>
<td>11</td>
<td>Low- vs. high-carbohydrate diet</td>
<td>17.6% vs. 17.3%</td>
<td>38% vs. 145%*</td>
</tr>
<tr>
<td></td>
<td>(n = 22)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Viljanen [29] (2009)</td>
<td>Uncontrolled</td>
<td>6</td>
<td>Balanced hypocaloric diet</td>
<td>11%</td>
<td>60%**</td>
</tr>
<tr>
<td></td>
<td>(n = 33)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haufe [40] (2011)</td>
<td>RCT</td>
<td>24</td>
<td>Low-fat vs. low-carbohydrate</td>
<td>16.5% vs. 17.5%</td>
<td>42% vs. 147%**</td>
</tr>
<tr>
<td></td>
<td>(n = 102)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

MRS, magnetic resonance imaging and spectroscopy; RCT, randomized, controlled trial.

n = based on number of completers.

*p < 0.05 and

**p < 0.01 when compared against baseline.

p values for between-groups comparisons are not statistically significant.
effective in reducing weight and excessive hepatic triglyceride accumulation.

Based on the above findings, it may be concluded that the potential impact of dietary composition is less important than the overall weight reduction achieved. Unfortunately, studies have been of short duration (≤6 months) and few have tried to assess the effect of dietary composition alone during weight-stable diets (summarized in Table 20.2). As can be observed in Table 20.2, isocaloric diets that do not affect overall caloric intake and weight have only a minor impact on hepatic steatosis (∼13–20%) [22,23], suggesting that weight loss per se may be the most important factor in ameliorating hepatic steatosis.

Physical activity only

Physical activity has also been suggested to be beneficial in patients with NAFLD and NASH [48]. In a large cohort of 813 patients, 54% of patients with biopsy-proven NAFLD were considered to be physically inactive when screened using a self-reported physical activity questionnaire [49]. Of note, the majority of patients with NAFLD reported that they did not engage at all in any physical activity. Though limited by the nature of the data collection (survey compared to direct observation), the above study suggests that increasing physical activity may play a role in the prevention or treatment of NAFLD.

Results among the many exercise studies in NAFLD are difficult to interpret due to heterogeneity in the type of exercise, baseline fitness levels, and amount of weight loss achieved. Moreover, few offer details about the changes in caloric content or dietary composition during the study. Like with dietary interventions, many of these trials have relied on weak surrogate biomarkers for the baseline diagnosis and follow-up of patients with NAFLD, such as liver aminotransferases and/or liver ultrasound [50–52].

Table 20.3 summarizes several studies that have used liver fat by MRS as the primary outcome of exercise-only interventions in patients with NAFLD. With some exceptions [53,54], all of them reported some reduction in hepatic steatosis after moderate exercise training [25–27,41,55]. Bacchi et al. [41] concluded that when aerobic and resistance exercise lead to a similar body weight reduction both are equally effective in reducing hepatic steatosis in patients with T2DM. In a recent study by Haus et al. [54] lack of weight reduction after 7 days of aerobic exercise resulted in no changes in total intrahepatic triglyceride content, emphasizing the importance of weight loss to achieve liver fat reduction. As can be appreciated in Table 20.3, most studies have been small and exercise alone (in the absence of weight reduction) was only moderately effective to reduce hepatic steatosis (∼10–20%) [25–27,54]. Patients who also lost weight had a greater reduction in liver fat content.

In a systematic review, Keating et al. [48] pooled data from 12 clinical trials involving 439 subjects. This work included an heterogeneous combination of studies that evaluated changes following physical activity on either liver aminotransferases, liver US, or hepatic fat content by MRS. Unfortunately, exercise intervention in most of the studies did not reach the minimal recommended exercise for the management of obesity and T2DM. While the effect on liver aminotransferases was inconsistent, the authors concluded that exercise alone had a small positive effect to decrease hepatic triglyceride accumulation in patients with NAFLD.

In summary, though the beneficial role of exercise per se has not been clearly demonstrated, there is no doubt that this intervention is beneficial in patients with NAFLD/NASH and/or T2DM, and that it should be recommended to patients in the management of NAFLD and NASH [13]. More work is clearly needed to establish the type, quality, and intensity of long-term exercise prescriptions in this population.

**Combined dietary intervention and exercise**

The available information on the role of combined lifestyle interventions (hypocaloric diets plus exercise) in patients with NAFLD is much more extensive, with several randomized, controlled trials showing significant benefit with this comprehensive approach [28,32,33,38,56–58]. Again, the overall reduction in liver fat reported in most of these trials has been

---

**Table 20.2. Studies assessing eucaloric low-fat vs. high-fat diets in patients with NAFLD**

<table>
<thead>
<tr>
<th>Author [ref.] (year)</th>
<th>Type of study</th>
<th>Duration (weeks)</th>
<th>Type of isocaloric diet</th>
<th>Weight loss</th>
<th>Change in LFAT by MRS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Westerbacka [22] (2005)</td>
<td>RCT-crossover (n = 10)</td>
<td>2</td>
<td>Low-fat vs. high-fat</td>
<td>NS</td>
<td>∼10% vs. 13.5%*</td>
</tr>
<tr>
<td>van Herpen [23] (2011)</td>
<td>RCT (n = 20)</td>
<td>3</td>
<td>Low-fat vs. high-fat</td>
<td>NS</td>
<td>∼13% vs. 11.7%*</td>
</tr>
<tr>
<td>Utzschneider [24] (2013)</td>
<td>RCT (n = 35)</td>
<td>4</td>
<td>Low-fat vs. high-fat</td>
<td>NS</td>
<td>∼20% vs. NS*</td>
</tr>
</tbody>
</table>

MRS, magnetic resonance imaging and spectroscopy; NS, not significant; RCT, randomized, controlled trial. n = based on number of completers. *p values < 0.05 for between groups comparisons. *p NS between groups comparisons.
Table 20.3 Studies assessing exercise-only interventions in patients with NAFLD

<table>
<thead>
<tr>
<th>Author [ref.] (year)</th>
<th>Type of study</th>
<th>Duration (weeks)</th>
<th>Exercise intervention (versus control)</th>
<th>Weight loss</th>
<th>Change in liver fat by MRS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shojaee-Moradie [53] (2007)</td>
<td>RCT (n = 17)</td>
<td>6</td>
<td>Aerobic (60–85% VO₂ max)</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Bonekamp [55] (2008)</td>
<td>RCT (n = 45)</td>
<td>24</td>
<td>Aerobic + resistance (80% max HR)</td>
<td>NS</td>
<td>↓2.5% ** (absolute)</td>
</tr>
<tr>
<td>Johnson [25] (2009)</td>
<td>RCT (n = 19)</td>
<td>4</td>
<td>Aerobic (50–70% VO₂ max)</td>
<td>NS</td>
<td>↓21% *</td>
</tr>
<tr>
<td>Hallsworth [26] (2011)</td>
<td>RCT (n = 19)</td>
<td>8</td>
<td>Resistance</td>
<td>NS</td>
<td>↓13% *</td>
</tr>
<tr>
<td>Sullivan [27] (2012)</td>
<td>RCT (n = 18)</td>
<td>16</td>
<td>Aerobic (45–55% VO₂ max)</td>
<td>NS</td>
<td>↓10% *</td>
</tr>
<tr>
<td>Bacchi [41] (2013)</td>
<td>RCT (n = 30)</td>
<td>16</td>
<td>Aerobic vs. resistance (60–65% max HR)</td>
<td>↓2.3% vs. ↓1.9% ^</td>
<td>↓32% vs. ↓26% ^</td>
</tr>
<tr>
<td>Haus [54] (2013)</td>
<td>RCT (n = 17)</td>
<td>1</td>
<td>Aerobic (80–85% max HR)</td>
<td>NS</td>
<td>NS</td>
</tr>
</tbody>
</table>

MRS, magnetic resonance imaging and spectroscopy; RCT, randomized, controlled trial. 
*n = based on number of completers.  
* p < 0.05 and  
** p < 0.01 when compared against control.  
^ p NS between groups comparisons.

strongly correlated with the amount of weight loss. Table 20.4 summarizes the studies that have examined the effect of lifestyle intervention on liver fat by MRS. As expected, those studies that achieved a lesser degree of weight loss [28,31] had a smaller reduction in liver fat content. Of note, studies that have compared hypocaloric diet versus hypocaloric diet plus exercise have failed to report a significant difference in the improvement of liver fat content with the combined strategy, but in both studies weight loss was about the same with both interventions [28,56]. Keating et al. [48] reached a similar conclusion in a meta-analysis of all available studies, in which exercise plus diet had no significant pooled effect size (ES) when compared to diet alone (ES: −0.05; 95% CI −0.38–0.27, p = 0.76).

Table 20.4 Studies assessing the effect of dietary plus exercise intervention on liver fat content by MRS

<table>
<thead>
<tr>
<th>Author (year)</th>
<th>Type of study (n)</th>
<th>Duration (weeks)</th>
<th>Main intervention</th>
<th>Weight loss</th>
<th>Change in liver fat by MRS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tamura [28] (2005)</td>
<td>RCT (n = 14)</td>
<td>2</td>
<td>Hypocaloric diet ± exercise Diet (−25% kcal)</td>
<td>↓1.5% vs. ↓2.3% ^</td>
<td>↓121% vs. ↓120% ^ ↓37% ^</td>
</tr>
<tr>
<td>Larson-Meyer [57] (2006)</td>
<td>RCT (n = 46)</td>
<td>24</td>
<td>Diet + exercise Diet (−15% WR)</td>
<td>↓10% ***</td>
<td>↓29% *** ↓40% ***</td>
</tr>
<tr>
<td>Kantartzis [31] (2008)</td>
<td>Uncontrolled (n = 50)</td>
<td>39</td>
<td>Hypocaloric diet + aerobic exercise (moderate)</td>
<td>↓3.5% **</td>
<td>↓35% **</td>
</tr>
<tr>
<td>Shah [56] (2009)</td>
<td>RCT (n = 18)</td>
<td>24</td>
<td>Hypocaloric diet ± exercise</td>
<td>↓9% vs. ↓10% ^</td>
<td>↓46% vs. ↓45% ^</td>
</tr>
<tr>
<td>Lazo [32] (2010)</td>
<td>RCT (n = 96)</td>
<td>48</td>
<td>Hypocaloric diet + exercise</td>
<td>↓18% ^</td>
<td>↓51% *</td>
</tr>
<tr>
<td>Wong [58] (2013)</td>
<td>RCT (n = 154)</td>
<td>48</td>
<td>Hypocaloric diet + exercise</td>
<td>↓18% **</td>
<td>↓55% **</td>
</tr>
</tbody>
</table>

MRS, magnetic resonance imaging and spectroscopy; RCT, randomized, controlled trial; WR, weight reduction. 
*n = based on number of completers.  
* p < 0.05 and  
** p < 0.01 when compared against control (vs. baseline for uncontrolled studies).  
^ p NS between groups comparisons.
An unresolved issue about many of these trials is whether small absolute changes in liver fat (even if of statistical significance) truly have clinical relevance. For instance, Lazo et al. [32] and Larson-Meyer et al. [57] reported apparently important reductions in liver fat content by MRS of 51% and ~35%, respectively, but that reflected only a slight absolute change in hepatic triglycerides within the normal range (note: NAFLD by MRS is defined as being >5.5%) translate into a true meaningful clinical benefit, or liver histologic improvement, in patients with NAFLD/NASH.

A few studies have performed liver biopsies before and after a weight loss intervention. As discussed earlier, histologic outcomes could be predicted by the magnitude of weight loss. In a 12-month uncontrolled study [59] using a standardized nutritional supplement combined with exercise counseling, failure to reduce weight was associated with no improvement in liver histology in the 15 (out of 23 enrolled) patients who completed the study. Consistent with the above, Eckard et al. [60] examined three different lifestyle interventions (low-fat diet or low-carbohydrate with moderate exercise or moderate exercise alone) compared to a standard care control group. Small study groups (11 to 12 per arm) and a high dropout rate (27%) limited the study conclusions. As in the previous study, failure to lose weight with diet and exercise (~1% weight loss) translated into no major histologic improvement compared to the control group. In contrast, Promrat et al. [33] found that 48 weeks of lifestyle intervention (diet plus moderate exercise), that significantly reduced weight when compared to controls (9.3% vs. 0.2%, p < 0.01), markedly improved the NAFLD activity score (NAS) (−2.4 vs. −1.4, p = 0.05). Of note, this was mainly due to an improvement in steatosis (−1.1 vs. −0.3, p = 0.02), as lobular inflammation and ballooning were similarly improved in both groups, while fibrosis was unchanged. However, when the data was re-analyzed for subjects who had lost ≥7% of body weight, lobular inflammation and ballooning were significantly improved (as well as steatosis), underscoring the importance of the magnitude of weight loss to histologic benefit.

Although there are no large, long-term studies to assess the role of lifestyle intervention in patients with NAFLD/NASH, the importance of this strategy in the management of patients with NAFLD cannot be overemphasized. Weight loss achieved through different strategies appears to be the most important factor to improve plasma aminotransferases, liver fat content, and histology. The role of lifestyle changes in the absence of weight reduction is less clear and needs further assessment in well-designed studies.

**Dietary intervention and weight-loss medications**

Because long-term compliance with lifestyle interventions remains a major challenge, several investigators have examined the role of pharmacologic interventions to enhance weight loss in patients with NAFLD/NASH. In a double-blind, randomized, controlled trial (RCT) by Zelber-Sagi et al. [61], 52 patients were randomized to lifestyle intervention ± orlistat for 6 months. After treatment there was no difference in weight loss between the two groups (−8% vs. −6%, p = 0.26). Histologic changes were similar in both groups. In another RCT in 50 obese patients with NASH, adding orlistat for 36 weeks to lifestyle intervention and vitamin E (800 UI daily) did not enhance weight loss (−8.3% vs. −6.0%, p = NS) or impact liver histology (primary endpoint) [62]. However, subjects who lost ≥5% of body weight markedly improved hepatic steatosis, while amelioration of liver necrosis and lobular inflammation required a body weight loss of at least 9%, suggesting that there is a lower threshold for steatosis resolution compared to other histologic parameters.

In summary, the role of the weight-loss medications tested to date in NAFLD is marginal. Large RCTs are much needed to assess the effectiveness of the recently FDA-approved weight management agents lorcaserin (Belviq®) and phentermine/topiramate extended-release (Qsymia®) in patients with NAFLD.

**Surgically induced dietary intervention**

The use of bariatric surgery as an alternative method to lose weight has become more popular in recent decades given the challenge of long-term weight loss. Different procedures are currently available and can be divided into three types: restrictive (reduction of gastric capacity), malabsorptive (bypasses a segment of the small bowel), and hybrid procedures (combination of the above) [63]. Currently, the most popular procedures are Roux-en-Y gastric bypass (RYGB; hybrid) and laparoscopic adjustable gastric banding (LAGB; restrictive). There is large evidence on the effects of bariatric surgery on glucose control [64], insulin sensitivity [65], and overall improvement of related comorbidities (hypertension, dyslipidemia, and so on) [64] as well as positive effects on long-term survival [66]. However, the impact of these procedures on liver histology in patients with NAFLD has not been so extensively addressed. Moreover, most of the existing studies are retrospective, uncontrolled or have a small sample size [67–78].

Prospective studies [69–74,78] specifically assessing the role of RYGB have reported overall positive results with improvement in steatosis, inflammation, ballooning, and to a lesser extent fibrosis, after 12–24 months of follow-up. In some reports, histologic improvement led to complete NASH resolution [69,71,73]. Of note, in all of these studies there was a very significant body weight reduction after surgery. The main drawbacks of these studies, however, have been their sample size (7 to 19 patients, except for one study [78] that included 70 patients) and the lack of a control group. In addition, a few studies have reported that some patients develop new fibrosis after the surgical intervention [68,71,72]. As a consequence of these findings, there have been some concerns regarding the effect of bariatric surgery on hepatic fibrosis. Several hypotheses have been suggested to explain the potential mechanisms responsible...
for worsening fibrosis in a subgroup of patients: rapid weight loss, lack of macro/micronutrients replacement or the direct effect of the procedure in a liver with pre-existing fibrosis. Of note, other series have not found any worsening of fibrosis in patients with NASH after bariatric surgery [69,73,74,78].

Regarding LAGB, although fewer studies have assessed its effect on histology in patients with NAFLD, studies are larger (n = 36 to 381) and many of longer duration (25–50 months) [75–77]. Similarly to RYGB, after major weight loss there is usually a significant improvement in steatosis, inflammation, and fibrosis. However, Mathurin et al. [77], in a large long-term follow-up study (381 patients followed for up to 5 years), found no change in hepatic inflammation and even a significant increase in fibrosis after one year, but without any further deterioration over time.

At present, there are no RCTs that have prospectively evaluated the best bariatric surgery approach for patients with NAFLD. It is also not well understood if the histologic changes after bariatric surgery in NASH are just the consequence of weight reduction or are secondary to other metabolic effects, such as changes in the gut microbiota or intestinal hormones. Further investigation is needed to determine the real long-term benefit of bariatric surgery in patients with NASH.

Pharmacologic agents

Statins

The role of statins in the primary and secondary prevention of CVD has been well established. Patients with NAFLD are believed to have increased cardiovascular risk [6] and are logical candidates for their long-term use. In accordance with their higher cardiovascular risk, the use of statins should be encouraged in patients with NAFLD but their use has remained controversial in such patients, particularly in the setting of elevated liver enzymes [5]. Recent practice guidelines on statin use in patients with NAFLD have clearly established that they are overall safe and that they do not carry a higher risk of liver toxicity [79,80].

In addition to their apparent safety in NAFLD, several small studies have suggested that statins may improve histology in NASH by means of their anti-inflammatory, antioxidant, or other pleiotropic properties [5]. Several studies reported that statins may decrease plasma aminotransferase concentration or hepatic steatosis by ultrasound [5,81,82]. However, it should be emphasized that these studies were overall of inadequate quality due to small sample size, uncontrolled design, short duration and/or lack of gold-standard endpoints, such as liver MRS or histology. In many of these trials effects were difficult to separate from weight loss, dietary changes or other lifestyle modifications during the study. In studies in which a liver biopsy was performed before and after statin treatment, Kimura et al. [83] (n = 43) and Rallidis et al. [84] (n = 5) found that liver inflammation improved, in contrast to reports from Georgescu et al. [85] (n = 10) and Ekstedt et al. [86] (n = 17). Beyond these discrepancies, there is agreement that hepatocyte ballooning and fibrosis do not improve with statin therapy. In the only randomized controlled trial in patients with biopsy-proven NASH, Nelson et al. [87] reported a nonsignificant trend towards a reduction in plasma aminotransferase concentration but no changes in liver histology. However, the study was rather small (n = 16) and of relatively short duration (12 months).

Taken together, the evidence showing a beneficial effect of statins in NAFLD or NASH is weak and awaits well-designed long-term studies. Their use should be promoted to prevent cardiovascular disease, but not as a means for histologic improvement.

Fibrates

Recent large randomized controlled fibrate trials in patients with T2DM have failed to meet the primary outcome (reduction of a major cardiovascular or coronary event) when combined with statins [88,89]. However, post-hoc analyses have shown that patients with high plasma triglycerides (>200 mg dL⁻¹) and low HDL-C concentration may benefit from combined statin and fibrate therapy [90]. Patients with NAFLD are characterized by this type of dyslipidemia, and therefore, are good candidates for combination therapy.

As discussed earlier for statins, there has been substantial interest on the role of fibrates in NAFLD. Fibrates have many biologic effects on hepatic lipid metabolism (most mediated by PPAR-γ effects) that appear to support their potential role in NAFLD. Fibrates have been reported to improve plasma aminotransferase levels or decrease hepatic steatosis, but overall results have been inconsistent and histology unchanged [5]. In a study by Athyros et al. [91], atorvastatin plus fenofibrate in patients with NAFLD had no effect beyond that of atorvastatin alone on plasma aminotransferase concentration or liver fat measured by ultrasound. In a well-designed RCT, two months of fenofibrate therapy had no effect on liver triglyceride accumulation when assessed by MRS [92].

In summary, fibrates are considered safe when needed for the prevention of cardiovascular disease in patients with NAFLD, but according to available evidence, they do not improve liver histology.

Ezetimibe

Ezetimibe has been evaluated in a few small studies in patients with NAFLD and atherogenic dyslipidemia. It has been reported to improve liver enzymes, inflammatory markers, and some histologic parameters [93,94]. In a randomized, single-blinded trial ezetimibe decreased liver fat by MRS but histologic outcomes were not examined [17]. However, these studies were small, open-label, or uncontrolled. Therefore, it is difficult to interpret whether these changes were the effect of ezetimibe therapy or just the consequence of weight reduction or lifestyle modification.
Pentoxifylline

Pentoxifylline is a nonselective phosphodiesterase inhibitor that reduces TNF-α and may affect other inflammatory pathways. Due to these anti-inflammatory effects, it has been considered as a potential treatment for NASH [95]. Although several studies have assessed the role of this drug in patients with NASH, just a couple of these were RCTs [95]. In one of such studies, Zein et al. [14] reported benefit on steatosis, inflammation, and fibrosis in 55 patients (46 with biopsy at the end of the study) randomized to pentoxifylline or placebo for one year. However, at baseline the two groups were not completely well matched, with patients in the pentoxifylline group having higher levels of AST/ALT and a trend towards worse histology. In another 12-month trial in 30 patients with NASH [15], pentoxifylline did not improve histology compared to placebo but decreased steatosis and ballooning against baseline. Once again, patients randomized to pentoxifylline had a worse baseline liver histology. Others studies assessing the role of this drug in NASH were either open-label, had a small sample size, or did not include histology or liver fat by MRS as endpoints [95]. Although the results with pentoxifylline are encouraging (especially the results of Zein et al. showing an improvement in fibrosis), much more work remains to be done to clearly understand the true role of this drug in NAFLD.

Omega-3 polyunsaturated fatty acids

Omega-3 polyunsaturated fatty acids (PUFAs) have been proposed as a potential therapy for liver steatosis [96]. They have been shown to activate PPARα receptors, which upregulate several genes involved in fatty acid oxidation and at the same time downregulate pro-inflammatory genes [97]. Moreover, PUFAs activate PPARγ, which also increases fat oxidation. Other important effects of PUFAs include reduction of endogenous lipid production and inhibition of hepatic glycolysis. Animal models have shown not only that low-PUFA diets promote steatosis and insulin resistance, but also that supplementing diet with PUFAs can prevent and even reverse hepatic steatosis [97]. Although several studies have examined the role of these compounds in humans, most of these studies have significant study design limitations [98,99], such as being open-label and lack of proper controls, or use as the primary outcome improvement of liver enzymes or liver fat by ultrasound. In the only RCT with PUFAs, Cussons et al. [100] reported a small, but statistically significant, reduction in hepatic steatosis by MRS. However, as discussed earlier, the meaning from a clinical standpoint of a small reduction in liver triglyceride content (from 10.2% to 8.4%) remains uncertain. Also, just one study assessed the effect of PUFAs on liver histology, but was small (n = 23) and uncontrolled [101]. Therefore, based on the available evidence it is difficult to make a final conclusion in regard to the role of PUFAs in NAFLD. Larger, well-designed RCTs are needed to understand their true effects in this disease.

Vitamin E

Vitamin E is an antioxidant believed to reduce hepatocyte oxidative stress in patients with NASH [102–105]. Several small studies of short duration in patients with NASH have reported inconsistent results [102–104,106–108]. In nondiabetic patients with biopsy-proven NASH vitamin E led to a 43% response in the primary histologic endpoint compared to 19% in the placebo group (p = 0.001). The primary histologic endpoint was an improvement in ≥2 grades in the NAS with at least one point improvement in hepatocellular ballooning and one point in either the steatosis or lobular inflammation score, with no worsening of fibrosis.

In summary, vitamin E may be beneficial in patients with NASH without T2DM because it appears safe and is relatively inexpensive, but its long-term efficacy (beyond 2 years) has not been established. Moreover, its use in patients with diabetes and NASH has never been studied and remains to be established.

Glucose-lowering agents

Metformin

Metformin is a biguanide that has been used worldwide for more than 50 years for the treatment of T2DM although its exact mechanism of action remains incompletely understood. It decreases hepatic glucose production and improves insulin sensitivity at the level of the liver, and to a lesser extent, muscle [109]. Metformin has been reported to reduce plasma aminotransferase levels in patients with NAFLD [104,110–112]. Early small studies suggested that metformin improved hepatic steatosis, necro-inflammation and/or fibrosis in patients with NASH [104,113,114], but not in more recent and better controlled clinical trials [115].

Taken together, although metformin is not recommended as a specific treatment for patients with NASH [13], it still has significant clinical value to control hyperglycemia and reduce cardiovascular risk in patients with prediabetes or T2DM and NASH [116].

Glucagon-like peptide 1 agonists

Gliclazide-like peptide 1 (GLP-1) analogues have become widely used in patients with T2DM and their beneficial effects have been quite consistent across studies in this population [117,118]. Although their role in NAFLD is more uncertain, recent studies have suggested a therapeutic potential for these pharmacologic agents in NAFLD. In animal models, exendin-4 has been reported to improve both liver function tests and histology [119,120]. Moreover, in vitro studies have shown that human hepatocytes express the GLP-1 receptor and that GLP-1 analogues suppress hepatic lipogenesis and pro-inflammatory mediators [121]. In humans with hepatic steatosis, open-label studies have shown that exenatide may improve liver enzymes and decrease steatosis when assessed by MRS [122,123], and even improve histology [124]. In a randomized, open-label trial by Sathyararayana et al. [125], the combination of pioglitazone
and exenatide improved transaminases and liver triglyceride accumulation by MRS beyond that of pioglitazone alone. In patients with T2DM treated with metformin, Jendle et al. [126] found that the GLP-1 analogue liraglutide added to metformin improved liver aminotransferase concentration and liver fat by computed tomography (CT) significantly more than the addition of glimepiride to the biguanide. Despite these promising results, the underlying mechanisms remain unclear and the relative contribution of direct GLP-1 agonism on the liver, versus weight loss or better glycemic control per se, demand further investigation in well-designed studies.

Thiazolidinediones (TZDs)

Thiazolidinediones (TZDs) are glucose-lowering agents that act as insulin sensitizers in humans. They are ligands for the transcription factor peroxisomal proliferator activated receptor-\(\gamma\) (PPAR-\(\gamma\)) that plays key roles in the regulation of metabolic homeostasis and inflammation. PPAR-\(\gamma\) is predominately expressed in adipose tissue, but is also present in the pancreas, liver, spleen, heart, and muscle [127]. In NASH, TZDs improve adipose tissue and hepatic insulin sensitivity, reduce subclinical inflammation, and restore liver histology [115,128].

Both pioglitazone and rosiglitazone have been investigated in NASH/NASH [129]. Most of the early studies were heterogeneous in terms of baseline patient characteristics and showed inconsistent clinical efficacy [103,112,130–132]. However, across these studies TZDs usually led to improved insulin sensitivity, reduced liver fat content, and normalization of plasma aminotransferase levels [103,130,133].

In patients with biopsy-proven NASH, the efficacy of TZDs, and especially of pioglitazone, has been examined in several studies (Table 20.5). The first RCT was reported in patients with prediabetes or T2DM and NASH [10]. In this population, pioglitazone (45 mg day\(^{-1}\)) significantly reduced insulin resistance at the level of the liver, adipose tissue and skeletal muscle, and improved hepatic steatosis, inflammation and hepatocellular ballooning when compared to placebo. In 73% of patients treated with pioglitazone, the NAFLD activity score improved compared to 24% in the placebo group (\(p<0.001\)).

A larger study later extended these findings to patients with NASH but without diabetes [11]. They randomized 247 subjects to vitamin E, pioglitazone, or placebo and found histologic improvement in liver steatosis and inflammation but not fibrosis after pioglitazone treatment. On the other hand, Ratziu et al. [132] evaluated the use of rosiglitazone in patients with NASH and reported more modest results, with only a \(\sim 20\)% reduction in hepatic steatosis on histology but no improvement in lobular inflammation, ballooning, or fibrosis.

Table 20.5 Histologic outcomes in randomized, placebo-controlled trials assessing the efficacy of pioglitazone in patients with NASH

<table>
<thead>
<tr>
<th>Author (year)</th>
<th>(n)</th>
<th>Duration (months)</th>
<th>Dose (mg)</th>
<th>Patients with diabetes</th>
<th>Patients with histologic improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Steatosis</td>
</tr>
<tr>
<td>Aithal [131] (2008)</td>
<td>74</td>
<td>12</td>
<td>30</td>
<td>No</td>
<td>NS</td>
</tr>
<tr>
<td>Cusi [134] (2013)</td>
<td>101</td>
<td>18</td>
<td>45</td>
<td>Yes</td>
<td>75%**</td>
</tr>
</tbody>
</table>

| Diabetes, type 2 diabetes mellitus; NS, not significant.  
| Study had a third arm with vitamin E 800 UI (see section on vitamin E).  
| \(p \leq 0.05\) and  
| \(**p < 0.01\) when compared against the control group.  
| Change not statistically significant compared to control group, but there was a statistically significant difference when change in mean scores was compared.  

## Confirmations and Considerations

Confirmation about TZDs’ long-term benefit is needed because the studies have been of relative short duration (6 to 24 months). In a preliminary report in patients with T2DM and prediabetes, Cusi et al. [134] have reported that after 18 months of treatment more subjects had improvement with pioglitazone compared to placebo for liver steatosis (75% vs. 29%, \(p<0.001\)), ballooning (55% vs. 26%, \(p<0.01\)), and inflammation (53% vs. 26%, \(p=0.01\)), as well as in the NAFLD activity score (NAS) (83% vs. 38%, \(p<0.001\)). Moreover, pioglitazone had a modest but significant effect on fibrosis compared to placebo (\(p=0.03\)) and baseline fibrosis stage (\(p<0.01\)).

With pioglitazone being the only TZD available in the United States and most countries, it must be kept in mind that while an effective drug for T2DM and NASH, caution must be exerted about its off-target effects (Figure 20.1) [135]. Thiazolidinediones have the potential to exacerbate congestive heart failure and promote water retention. Moreover, its use has been associated with a greater risk of osteoporosis and bladder cancer. In a meta-analysis of 19 trials in 16,390 patients with T2DM [136], the rate of congestive heart failure was noted to be slightly higher in pioglitazone users versus control patients using other oral agents for diabetes (2.3% vs. 1.8%, HR 1.41; 95% CI 1.14–1.76; \(p=0.002\)). However, there was also a significant reduction in the combined outcome of death, myocardial infarction, or stroke (\(p=0.005\)). Pioglitazone has been associated with bladder cancer, and currently the FDA recommends avoiding it if active bladder cancer is present, and use with caution if there is a prior history of the disease. However, a recent 8-year interim analysis (from January 1, 1997 to December 31, 2010) reported a trend for...
the risk of bladder cancer to be less over time, with hazard ratios (HR) that were nonsignificant in fully adjusted models that included key variables such as tobacco use, duration of pioglitazone therapy, or cumulative pioglitazone dose (Clinical Trial NCT01637935). For example, duration of therapy with pioglitazone for more than 2 years (HR = 1.4; 95% CI 1.03–2.0) or 4 years (HR = 1.62; 95% CI 0.96–2.74) in the 5-year analysis was higher than the 8-year analysis (HR = 1.30; CI 0.91–1.86, NS). Similarly, cumulative doses of pioglitazone >28,000 mg in the 5-year analysis (HR = 1.43; CI 0.96–2.12) did not continue to increase the risk of bladder cancer in the 8-year analysis (HR = 1.25; CI 0.91–1.74, NS). Recent guidelines from the American Association for the Study of Liver Diseases, American College of Gastroenterology, and the American Gastroenterological Association [13] recommend the use of pioglitazone as a valid alternative for the management of patients with biopsy-proven NASH in patients with diabetes. However, there is clearly a need for more studies to evaluate its long-term efficacy and safety. Careful patient selection is crucial in order to maximize benefits while reducing the risks associated with pioglitazone use, as recently reviewed elsewhere [135].

**Obeticholic acid (OCA)**

Obeticholic acid, a derivative of the primary human bile acid chenodeoxycholic acid, is an agonist of the farnesoid X receptor (FXR: a key nuclear receptor that regulates hepatic glucose and lipid metabolism). In a small study [137], OCA was found to only marginally ameliorate insulin resistance. Improvements in insulin sensitivity and liver aminotransferases only occurred at doses of 25 mg, but not at higher doses (50 mg). The same was true for plasma markers of inflammation and fibrosis. However, in early 2014, a RCT of OCA (25 mg day$^{-1}$) versus placebo in patients with biopsy-proven NASH was stopped early during an interim analysis because of significant positive results on liver histology. In data from about half of the 283 randomized
patients, it was concluded that OCA led to a highly significant improvement in histology according to an intention-to-treat analysis. However, long-term data is still needed to fully assess its safety and efficacy.

**Future directions**

Increased awareness among healthcare providers will be needed for the early diagnosis and optimal management of the patient with NAFLD/NASH. Patient care will entail going beyond liver-specific complications to embrace a comprehensive strategy that ameliorates obesity-induced lipotoxicity while addressing associated comorbidities, such as T2DM or CVD. Reminding ourselves that the condition is really that of a chronic “lipotoxic liver disease” helps highlight the crosstalk between dysregulated adipose tissue and the liver, be more aware of its chronic nature and the need for a long-term management plan. It is likely that diagnosis accuracy will improve in the near future with a combination of more specific plasma biomarkers, advances in genetic testing and better metabolic profiling (i.e., metabolomics approaches).

Earlier and more accurate disease staging will allow targeting therapy earlier and direct it only at patients with a greater risk of disease progression. Therapies that relieve the liver from systemic lipotoxicity by targeting adipose tissue insulin resistance and inflammation (e.g. significant weight loss, exercise, and/or pioglitazone) hold reasonable chances of long-term success. However, new and more effective approaches will be needed. One can envision the need for combining agents that target different defects, much like what we do today for T2DM or hypertension. Multiple agents are undergoing testing in RCTs for the treatment of NASH (see https://clinicaltrials.gov/). The different approaches proposed include “hepatoprotective” agents such as silymarin (milk thistle) or the antibiotic rifaximin. They are believed to improve the deleterious gut flora present in obesity and NAFLD and mediate positive changes in insulin sensitivity. Novel insulin-sensitizers such as resveratrol (an antioxidant), fenretinide (antagonize insulin action), or the new adenosine monophosphate-activated protein kinase (AMPK) activators (olitipraz) that improve insulin sensitivity and fatty acid synthesis, are undergoing investigation. Other approaches include the antioxidant S-adenosyl-L-methionine (SAMe), aimed at reducing lipid peroxidation and secondary cellular injury.

Until we find the optimal long-term lifestyle and pharmacologic approach for NAFLD in T2DM, based on the evidence reviewed, the clinician must remain alert and offer an integrated management plan to the increasing number of patients with this chronic and relentless liver condition.

**References**


83 Kimura Y, Hyogo H, Yamagishi S, et al.: Atorvastatin decreases serum levels of advanced glycation endproducts (AGEs) in nonalcoholic steatohepatitis (NASH) patients with dyslipidemia: clinical


100 Cussons AJ, Watts GF, Mori TA, Stuckey BG: Omega-3 fatty acid supplementation decreases liver fat content in polycystic ovary syndrome: a randomized controlled trial employing proton magnetic resonance spectroscopy. Journal of Clinical Endocrinology & Metabolism 2009;94:3842–3848.


Treatment of nonalcoholic fatty liver disease (NAFLD) and nonalcoholic steatohepatitis (NASH)


126 Jendle J, Nauck MA, Matthews DR, et al.: Weight loss with liraglutide, a once-daily human glucagon-like peptide-1 analogue for type 2 diabetes treatment as monotherapy or added to metformin, is primarily as a result of a reduction in fat tissue. Diabetes, Obesity and Metabolism 2009;11:1163–1172.


CHAPTER 21
Type 2 diabetes and cancer

Emily Jane Gallagher1, Ruslan Novosyadlyy2 and Derek LeRoith1
1 Division of Endocrinology, Diabetes and Bone Diseases, Icahn School of Medicine at Mount Sinai (ISMMS), New York, NY, USA
2 Department of Oncology—Translational Medicine, ImClone Systems, a wholly-owned subsidiary of Eli Lilly and Company, Alexandria Center for Life Science, New York, NY, USA

Key points

• Epidemiology demonstrates an association between obesity, type 2 diabetes and increased cancer risk and cancer-related mortality.
• Numerous potential factors may be causal such as insulin, IGFs, glucose, lipids, adipokines, and inflammatory cytokines.
• Mouse models are a valuable tool to study potential causal factors.

Introduction

An association between hyperglycemia, diabetes, and cancer has been recognized for many years. Epidemiologists first noted the association between diabetes and cancer in the early part of the twentieth century, while the association between hyperglycemia and cancer was reported in 1885. At that time, in Europe and North America life expectancy was improving, rates of over-nutrition and under-exercise were increasing, there was a rise in the percentage of people that were overweight, and the incidence of diabetes began to climb [1]. The epidemiologists and statisticians of the time examined the association between diabetes and cancer in different populations to determine if there was truly a relationship between these conditions, or if the association was purely related to increased longevity. Their studies found that although both conditions increased with age, diabetes appeared to increase the risk of cancer, independent of age. Meanwhile, physicians had observed differences in glucose homeostasis in individuals with cancer, and researchers such as Otto Warburg were studying the distinct metabolic properties of cancer cells, namely the generation of energy by the fermentation of glucose [2]. Since these early discoveries, a wealth of research has been conducted in the field of diabetes and cancer. Today, in the setting of the global diabetes and obesity epidemics, understanding the epidemiologic links between diabetes and cancer and determining the mechanisms through which these conditions are linked has become a priority for research, in order to prevent and treat cancers that are more likely to occur in those with diabetes.

Epidemiology

Epidemiologic studies continued through the twentieth century, with reports of a link between diabetes and the development of pancreatic, endometrial, breast cancer, and hematologic malignancies. However, until the 1990s much of the epidemiologic findings were inconsistent due to many confounding risk factors that were not accounted for in studies, such as obesity, smoking, and hormonal therapy as well as differences in study design. From the 1990s to present, large prospective cohort studies have been conducted, such as the Cancer Prevention Study II (CPSII), the Physicians’ Health Study and the Women’s Health Initiative (WHI) in the USA, and the European Prospective Study into Cancer. In these studies individuals were recruited, data including the presence or absence of diabetes, body mass index (BMI), smoking status, were collected along with serum samples. Participants were then followed for a number of years to determine whether they developed cancer or died from cancer. These studies reported an increased incidence of cancer in individuals with diabetes, and an increased mortality in those with diabetes who develop cancer.

Although many of these studies did not specifically record diabetes type, in the US of adults with diabetes, 90–95% have type 2 diabetes (T2DM). Therefore, at least 90% of the populations in these studies would have T2DM. In contrast to T2DM, the studies into the risk associated with type 1 diabetes (T1DM) and cancer have been largely inconsistent. One major potential reason for variable results is that the majority of studies use either insulin therapy or age cut-offs to define T1DM. These age cut-offs vary widely from under 18 years of age to under 40...
years of age; therefore, individuals with type 2 diabetes are very likely to be included in some of these studies. In addition, these studies are a mixture of cohort and case-control studies, which may also contribute to the variable results. With these limitations in mind, individual studies have reported an increased risk of pancreatic, liver, mouth and pharyngeal, stomach, skin and ovarian cancer, as well as leukemia in those with T1DM. The types of cancers that are associated with T1DM are not consistently reported between the studies, and some studies report no association between diabetes and cancer [3].

While the overall risk of cancer is greater in those with T2DM, several studies and meta-analyses have examined which cancers are specifically affected by diabetes. While the earliest epidemiologic studies demonstrated a significantly increased risk of pancreatic cancer in those with diabetes, it was later noted that there was a close temporal relationship between the onset of diabetes and the diagnosis of pancreatic cancer; in fact the development of hyperglycemia and diabetes may be a manifestation of undiagnosed pancreatic cancer, this effect is known as “reverse-causation.” More recent studies have excluded individuals who develop cancers within 5 years of the start of the study in an attempt to avoid this effect and these studies have demonstrated an increase in the risk of pancreatic among other cancers in those with T2DM [4]. The results of the meta-analyses examining specific cancer sites and T2DM are summarized in Table 21.1. Individuals with diabetes have approximately twice the risk of hepatocellular, endometrial, and pancreatic cancer than those without diabetes [5–7]. The risk of biliary tract and renal cancer appear to be increased by 40% and bladder cancer by approximately 20% in those with diabetes [8–10], while the risk of breast cancer is increased by 20% in women and colorectal cancer is increased by 30% in both sexes [11,12]. The risk of esophageal cancer may be increased in men with diabetes, while the risk of certain brain cancers and gastric cancer may be increased in women with diabetes [13–15]. The most notable exception to the general increased risk of cancer is with prostate cancer; epidemiologic studies and meta-analyses have reported a decreased risk of developing prostate cancer in men with diabetes [16,17]. However, in some studies for prostate cancer, similar to those in breast and colorectal cancer, diabetes has been associated with a greater mortality risk, a greater risk of treatment failure, and higher recurrence rate. The majority of studies have been conducted in the US and Europe; however, the results are not necessarily applicable to other populations (Asian, African, and Hispanic) where the prevalence of diabetes, obesity, and different types of cancer varies (e.g. hepatocellular and esophageal cancer is more common in Asia, and other factors may outweigh any increased risk from diabetes). Studies are now being conducted in different racial and ethnic groups in different countries to determine whether the risk associated with diabetes is applicable worldwide. As set out in Table 21.1, meta-analyses have shown that individuals from Asia with a history of diabetes have no increased risk of esophageal cancer (in contrast to the European population), while similar to other groups they do have an increased risk of hepatocellular cancer, and in contrast to other ethnic groups may have an increased risk of prostate cancer [7,14,17]. Although there is substantial evidence supporting the link between diabetes and the more common epithelial cancers, there is less information available on the less common cancers [4]. These cancers are more difficult to study in large epidemiologic studies, due to their lower incidence in the general population and therefore longer follow-up and larger studies are needed to determine whether diabetes increases the risk of some of the less common cancers.

<table>
<thead>
<tr>
<th>Cancer site</th>
<th>Overall relative risk (95% CI)</th>
<th>Male relative risk (95% CI)</th>
<th>Female relative risk (95% CI)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liver</td>
<td>2.31 (1.87–2.84)</td>
<td>2.03 (1.58–2.62)</td>
<td>1.91 (1.22–2.99)</td>
<td>[7]</td>
</tr>
<tr>
<td></td>
<td>2.00 (1.46–2.75)a</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Endometrium</td>
<td>N/A</td>
<td>1.70 (1.55–1.87)</td>
<td>1.60 (1.43–1.77)</td>
<td>[6]</td>
</tr>
<tr>
<td>Pancreas</td>
<td>1.83 (1.38–2.43)b</td>
<td>1.31 (1.17–1.47)</td>
<td>1.29 (1.13–1.46)</td>
<td>[5]</td>
</tr>
<tr>
<td>Biliary tract</td>
<td>1.43 (1.18–1.72)</td>
<td>1.26 (1.06–1.49)</td>
<td>1.70 (1.49–1.97)</td>
<td>[10]</td>
</tr>
<tr>
<td>Kidney</td>
<td>1.42 (1.06–1.91)</td>
<td>1.29 (1.15–1.44)</td>
<td>1.33 (1.23–1.44)</td>
<td>[11]</td>
</tr>
<tr>
<td>Colorectal</td>
<td>1.30 (1.20–1.40)</td>
<td>1.28 (1.10–1.49)</td>
<td>1.07 (0.71–1.62)</td>
<td>[14]</td>
</tr>
<tr>
<td>Esophagus</td>
<td>1.30 (1.12–1.50)</td>
<td>1.26 (0.91–1.75)b</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bladder</td>
<td>1.24 (1.08–1.42)</td>
<td>Not reported</td>
<td>Not reported</td>
<td>[9]</td>
</tr>
<tr>
<td>Breast</td>
<td>Not reported</td>
<td>1.29 (0.99–1.67)</td>
<td>1.20 (1.13–1.29)</td>
<td>[12]</td>
</tr>
<tr>
<td>Brain</td>
<td>1.12 (0.89–1.42)</td>
<td>1.02 (0.94–1.12)</td>
<td>1.24 (1.03–1.50)</td>
<td>[13]</td>
</tr>
<tr>
<td>Stomach</td>
<td>1.09 (0.98–1.22)</td>
<td>1.04 (0.94–1.15)</td>
<td>1.18 (1.01–1.39)</td>
<td>[15]</td>
</tr>
<tr>
<td>Prostate</td>
<td>0.86 (0.80–0.92)</td>
<td>0.86 (0.80–0.92)</td>
<td>1.31 (1.12–1.54)a</td>
<td>[17]</td>
</tr>
</tbody>
</table>

aAsian population.
bDiabetes duration of at least 5 years before the diagnosis of pancreatic cancer.
In the CPS II study of 1,053,831 individuals after 26 years of follow-up, increases in mortality from less common cancers such as male breast cancer (fourfold increased risk) and oral and pharyngeal cancers (40% increased risk) were observed in those with diabetes. It should be noted that many studies report an increased risk of mortality in patients with cancer who have diabetes, but any individual with diabetes has almost doubled the age-adjusted mortality of someone without diabetes. Therefore, whether diabetes increases cancer-specific mortality because of more aggressive cancer growth and spread, or whether the increased mortality is merely a reflection of the overall increased mortality seen with diabetes remains unclear [4].

**Links between diabetes and cancer**

**Common risk factors**

Epidemiologic studies demonstrate that those with diabetes are more likely to develop cancer, but the mechanisms through which this may occur are incompletely understood. The American Diabetes Association and the American Cancer Society published a consensus report in 2010, with the aim of examining the knowledge regarding the association between diabetes and cancer, exploring the risk factors for both conditions, to examine their possible biologic links and to determine whether certain treatments for diabetes modify cancer risk [4]. In the report they identified certain risk factors that increase the chance of developing both conditions. These were listed as “non-modifiable” and “modifiable” risk factors that are common to diabetes and cancer and may explain in part the increased risk of cancer. Non-modifiable risk factors include age, sex, and race/ethnicity. While modifiable risk factors include overweight/obesity, smoking, alcohol intake, and physical activity and diet (Table 21.2). In addition there are biologic factors that may be common to the two conditions. With advancing age the risk of both T2DM and cancer increase: 78% of newly diagnosed cancers occur in those over 55 years of age, the prevalence of diabetes also increases with age, from 10.8% in those aged 40–59 years, but increases to 23.8% in those aged 60 years or more. Men have a slightly higher risk of developing diabetes than women, additionally, with the exception of certain sex-specific or almost sex-specific cancers (e.g. prostate, testicular, cervical, endometrial, and breast cancer), men are also more likely than women to develop cancer. Certain racial groups (e.g. African Americans) are more likely to develop cancer and diabetes than other racial groups. Overweight and obesity have been associated with cancers for many years. Large epidemiologic studies, including the Cancer Prevention Study II in the USA and the Million Women’s Study in the UK reported that for both men and women an increased body mass index (BMI) was associated with a greater risk of developing colorectal, esophageal, pancreatic, and kidney cancer, non-Hodgkin’s lymphoma and multiple myeloma; additionally they reported an increased risk of endometrial cancer, ovarian cancer, and postmenopausal breast cancer in women. In fact the results of the CPS II study suggest that 14% of cancer deaths in men and 20% in women could be attributed to obesity. Currently over a third of Americans are obese. Overweight and obesity also increase the risk of developing T2DM. A number of studies have now shown that smoking is an independent risk factor for developing T2DM and is also a risk factor for multiple cancers. Moderate to high alcohol consumption is known to increase the risk of certain cancers and in certain populations, such as in lean Japanese individuals moderate to high alcohol intake is also a risk factor for the development of diabetes [18]. A sedentary lifestyle is associated with an increase in many chronic diseases, amongst them are diabetes and cancer. Finally, consuming a diet rich in fruits and vegetables has been shown in studies to be protective against the development of T2DM and cancer.

**Biologic links**

Identifying the increased risk of cancer in those with T2DM or obesity let to the hypothesis that there must be common biologic links between these conditions that increase the risk of cancer. Additionally, recent studies have shown that those with the metabolic syndrome, a syndrome on the continuum between obesity and T2DM also have an increased risk of cancer. Physiologic changes that may occur in individuals with obesity, the metabolic syndrome and T2DM and may contribute to cancer development include hyperglycemia, insulin resistance and hyperinsulinemia, increased insulin-like growth factor-I (IGF-I) levels, dyslipidemia, visceral adiposity and increased inflammatory cytokines, altered levels of circulating adipokines, and altered circulating and tissue levels of estrogens (Table 21.3). Epidemiologic and animal studies have been conducted to examine these individual biologic factors on cancer development.

<table>
<thead>
<tr>
<th>Table 21.2</th>
<th>Risk factors that increase the chance of developing type 2 diabetes or cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Non-modifiable</strong></td>
<td><strong>Modifiable</strong></td>
</tr>
<tr>
<td>Age</td>
<td>Overweight/Obesity</td>
</tr>
<tr>
<td>Sex</td>
<td>Smoking</td>
</tr>
<tr>
<td>Race</td>
<td>Alcohol</td>
</tr>
<tr>
<td></td>
<td>Physical activity</td>
</tr>
<tr>
<td></td>
<td>Diet</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 21.3</th>
<th>Potential biologic links between type 2 diabetes and cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Hyperglycemia</td>
<td>• Hyperinsulinemia</td>
</tr>
<tr>
<td>• IGF-I</td>
<td>• Dyslipidemia</td>
</tr>
<tr>
<td>• Inflammatory cytokines</td>
<td>• Adipokines</td>
</tr>
<tr>
<td>• Circulating and tissue estrogen levels</td>
<td></td>
</tr>
</tbody>
</table>
Hyperglycemia

It is well known that cancers take up glucose, a phenomenon that is exploited in the visualization of fluorodeoxyglucose (FDG) uptake by positron emission tomography (PET) for the detection of tumors and metastases. Some epidemiologic studies have examined how glucose levels alter cancer risk. A European study of individuals with the metabolic syndrome (The Me-Can study) found that higher glucose levels were associated with an increased risk of liver, gallbladder, respiratory, thyroid cancer, and multiple myeloma in men, as well as pancreatic, bladder, endometrial, cervical, and stomach cancer in women. However, in a meta-analysis of prospective diabetes studies (VADT, ACCORD, UKPDS33, UKPDS34), complications were compared in patients receiving intensive glucose control and A1c values of 6.9%, 6.4%, 7%, and 7.9% with those receiving standard diabetes control and A1c values of 8.4%, 7.5%, 7.9%, and 8.5%, respectively. They determined that there was no difference in cancer outcome between the standard and intensively controlled groups. These studies were not designed to examine cancer incidence as a primary outcome; some of the treatments for diabetes may potentially have increased the risk. Additionally, all of the patients in these studies had a diagnosis of diabetes, and therefore were all at potentially increased risk of cancer as well as having abnormal glucose levels, compared to the normal nondiabetic population. Other prospective studies on over 60,000 individuals in the Vasterbotten Intervention Project in Sweden and the Hong Kong diabetes registry did report an increased risk of cancer in those with hyperglycemia. Thus, human studies assessing the effect of hyperglycemia on tumor growth are inconclusive due to confounding biologic and diabetes treatment effects.

Insulin and IGF-I

In many of the previously mentioned prospective epidemiologic studies, such as the WHI, the Physicians’ Health Study and the EPIC studies, serum samples were stored at the time of recruitment and these samples were subsequently analyzed to determine whether there was an association between a specific biomarker (e.g. endogenous insulin levels) and the risk of certain cancers. In the WHI study insulin levels were measured in the study population and then the population was divided into quartiles based on their endogenous insulin levels. Comparing those with insulin levels in the highest quartile with the lowest quartile, they found that women with higher insulin levels who did not take hormone replacement had a greater risk of breast cancer. The Nurses’ Health Study II, which consisted mostly of premenopausal women, did not find an association between insulin and premenopausal breast cancer. C-peptide is used in some studies as a marker of endogenous insulin secretion. Higher C-peptide levels have also been associated with postmenopausal breast cancer. In the Physicians’ Health Study and the Nurses’ Health Study higher levels of C-peptide were associated with an increased risk of colorectal cancer. Subsequent studies have also reported that high plasma C-peptide levels are associated with colonic adenomas, suggesting that hyperinsulinemia may act as an early etiological factor in the development of colon cancer. A meta-analysis of studies reported that individuals who develop colorectal and pancreatic cancer have increased prediagnostic insulin/C-peptide levels. Interestingly, the Physicians’ Health Study also found that high C-peptide levels were associated with an increased risk of prostate cancer and prostate cancer mortality. In a study from Finland, fasting insulin concentrations were higher in men who subsequently developed prostate cancer, an association that was independent of glucose levels. Some variability in results has been observed in studies on insulin/C-peptide and cancer risk and data are only available on a few cancer types. Many of these studies did not collect serum for insulin and C-peptide in fasting conditions, and these samples were collected and stored for many years long before analysis and before the development of cancer, therefore these may contribute to some variability between the epidemiologic studies.

Hyperinsulinemia directly increases hepatic IGF-I expression via the IR and leads indirectly to increased hepatic IGF-I production, by increasing hepatic growth hormone receptor levels. Growth hormone then acting through the growth hormone receptors stimulates IGF-I expression. High normal levels of circulating IGF-I are associated with increased risk of certain cancers including colorectal, breast, and prostate cancer in prospective epidemiologic studies and meta-analyses [19–21]. No association has been found between circulating IGF-I and lung cancer.

Dyslipidemia

The relationship between cholesterol and cancer risk has been difficult to discern from epidemiologic studies. Individuals with the metabolic syndrome and T2DM frequently have a dyslipidemic profile with elevated triglycerides and decreased high-density lipoprotein cholesterol (HDL-C). In the Metabolic Syndrome Cancer (Me-Can) European cohort elevated triglyceride levels were associated with colon, respiratory, kidney, thyroid cancers, and melanoma in men and respiratory, cervical, and nonmelanoma skin cancers in women [22]. An inverse relationship was found between elevated HDL-C and incident cancer risk in a meta-analysis of 24 randomized control trials, with inverse relationships also being found for both prostate and breast cancer risk. A recent Danish population study reported that the use of statins before the diagnosis of cancer was associated with reduced cancer-related mortality compared to nonusers of statins [23]. There is a relative dearth of studies examining the association between cholesterol and cancer, as elevated cholesterol is associated with such significant increases in cardiovascular mortality that many individuals died of cardiovascular disease before developing cancer. Additionally, caution should be applied when interpreting studies on cholesterol and cancer as low cholesterol levels are associated with chronic diseases, thus individuals with metastatic lung cancer may have low cholesterol levels, which
may be a cause of the chronic illness rather than low cholesterol levels being associated with metastatic lung cancer. Further epidemiologic studies in the area of cholesterol and cancer are necessary to determine whether a true association between these conditions exists.

**Inflammatory cytokines and adipokines**

Abdominal obesity is considered a state of chronic inflammation and this inflammation is often thought to contribute to the insulin resistance that leads to the development of the metabolic syndrome and T2DM. Macrophages found in adipose tissue secrete inflammatory cytokines including interleukin-6 (IL-6) and tumor necrosis factor alpha (TNF-α) and these cytokines have also been implicated in the development of numerous cancers. Examination of serum TNF-α and IL-6 levels in the EPIC cohort have found that higher circulating levels of these cytokines were associated with a significantly increased risk of endometrial cancer; however, adjusting for body mass index resulted in a loss of a significant association between IL-6 and endometrial cancer. This should not be considered surprising as the increase in inflammatory cytokines is thought to be associated with increased adiposity.

The circulating levels of the adipokine leptin increase in the setting of obesity, the metabolic syndrome, and T2DM, while circulating levels of adiponectin decrease. High leptin levels and low adiponectin levels have been associated with certain cancers, although epidemiologic studies have shown inconsistent results. In a cohort from the WHI, low adiponectin and high leptin levels were associated with an increased risk of colorectal cancers; however, the effect of low adiponectin was not independent of insulin. Low adiponectin has also been associated with an increased risk of breast, renal, and prostate cancer as well as multiple myeloma in some studies. Although leptin is hypothesized to increase tumor growth, the Physicians’ Health Study found no association between leptin and prostate cancer; no association has been found between leptin and renal cancer or multiple myeloma in the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial, and the Nurses’ Health Study found that leptin levels were inversely associated with premenopausal breast cancer risk. Overall, these epidemiologic studies suggest that both inflammatory cytokines and adipokines may be playing a role in the development of cancers in individuals with obesity, the metabolic syndrome, and T2DM. Whether these effects are independent of insulin remains to be determined.

**Altered estrogen levels in T2DM and cancer**

Obese postmenopausal women are known to have higher circulating levels of estrogen than their lean counterparts due to increased aromatase activity in adipose tissue. Additionally, obese insulin-resistant women have decreased hepatic production of sex hormone-binding globulin (SHBG) leading to more bioavailable estrogens in the circulation. Recent studies have shown that in the breast tissue from obese individuals, aromatase expression is increased approximately twofold, compared to the expression level in the breast tissue of normal weight women [24,25]. This increased expression may lead to higher estrogen concentrations in the breasts of obese women and may drive the growth of tumors that are hormone responsive, such as common postmenopausal breast cancers.

Overall, epidemiologic studies have demonstrated a consistent increase in the risk of many epithelial cancers in those with T2DM. Many risk factors have been identified that are associated with an increased risk of both T2DM and cancer development. Biologic links that are known to be associated with obesity, the metabolic syndrome, and T2DM have been identified, including hyperinsulinemia, increased IGF-I, dyslipidemia, inflammatory cytokines, adipokines and estrogens, and the association between these factors and cancer have been explored in epidemiologic studies. However, the current epidemiologic studies cannot directly determine whether these factors can actually cause cancer or increase tumor growth, or the mechanisms through which they may exert these effects. Therefore, animal models have been created in order to understand the links between diabetes and cancer.

**Animal studies**

**In experimental studies, type 1 and type 2 diabetes have a different impact on tumor growth**

A few decades ago it was shown that chemical destruction of pancreatic β cells with the subsequent development of T1DM suppresses tumor growth, whereas administration of exogenous insulin stimulates malignant growth in tumor-bearing animals with T1DM [26]. In addition, intraportal transplantation of pancreatic islets into rats with T1DM results in the development of an insulin-enriched microenvironment promoting hepatocellular tumor growth. Furthermore, in a model of pancreatic islet transplantation it has been shown that insulin-driven hepatocarcinogenesis is mediated by the PI3K/Akt/mTOR pathway and can be effectively inhibited by NVP-BEZ235, a dual PI3K/mTOR inhibitor. Taken together, the results of these studies indicate that (i) insulin is capable of stimulating tumor growth, (ii) T1DM is associated with a suppressed tumor growth, and (iii) hyperglycemia per se does not promote tumor growth, at least in the setting of insulinopenia.

The relationship between T2DM and cancer is much more complex. Most individuals with T2DM are overweight and/or obese, and each of the individual diabetes- and adipose tissue-related factors (hyperinsulinemia, hyperglycemia, hyperlipidemia, adipokines, cytokines, chemokines, and growth factors) is capable of stimulating tumor growth.

Most experimental studies investigating the link between T2DM and cancer were performed in rodent (primarily mouse) models of genetically induced and diet-induced obesity and T2DM. However, a lot of insight into the mechanisms that link T2DM and cancer was obtained from cancer studies conducted in models of obesity-independent T2DM.
Genetically induced obesity and cancer

Under experimental conditions, obesity in rodents is frequently induced by a genetic disruption of leptin signaling. Leptin interacts with a cognate receptor in hypothalamic neurons to decrease appetite and regulate energy homeostasis and thermogenesis. Rodent models of genetically induced obesity have been extensively used to study the relationship between obesity and/or insulin resistance and cancer.

The spontaneously occurring lethal yellow, or yellow agouti (Aγ), mouse strain is one of the first rodent models in which an association between increased body weight and a higher susceptibility to tumor development has been noted [27]. Later, a few new yellow mutations have been identified, including the viable yellow (Aγγ) mutation, which in contrast to the Aγ mutation, is not lethal when inherited homozygously. The yellow mice (Aγ, Aγγ) have a multifaceted phenotype which includes obesity, insulin resistance, enhanced linear growth, and susceptibility to tumor formation. The obese phenotype in the yellow mouse is caused by expression of abnormal agouti protein resulting in blockade of hypothalamic melanocortin-4 receptors, which are involved in the regulation of feeding behavior and energy homeostasis. However, the increased tumor growth in these mice should be interpreted with caution as several lines of evidence indicate that ectopic expression of agouti is capable of promoting chemically induced carcinogenesis independent of obesity and hyperinsulinemia [28].

The Lepob/Lepob (ob/ob) mice bear a naturally occurring autosomal recessive mutation in the obese (ob) gene encoding leptin [29]. ob/ob mice demonstrate markedly increased adiposity, hyperglycemia, and hyperinsulinemia.

Tumor growth in ob/ob mice was investigated after subcutaneous inoculation of murine melanoma B16 cells into female ob/ob and lean mice. In three different studies, primary tumor growth was unaffected, retarded, or enhanced in leptin-deficient mice compared to the lean controls, whereas tumor metastasis was suppressed in ob/ob mice. In contrast, intravenous inoculation of B16 cells in ob/ob mice was associated with an extensive formation of pulmonary metastasis. In vitro, serum obtained from ob/ob mice induces a mesenchymal phenotype and increases invasive potential of B16 cells which may explain a robust effect on pulmonary metastasis rather than primary tumor growth described earlier. A recent study also reported that the phenotype observed in ob/ob mice does not affect the development of carcinogen-induced skin papillomas and spontaneous p53-deficient malignancies (mostly lymphomas).

ob/ob mice, however, exhibited an increased incidence of preneoplastic lesions in the colon compared to the lean controls when treated with azoxymethane (AOM) and N-methyl-N-nitrosourea (NMU), two commonly used chemical carcinogens. When ob/ob mice were crossed with immunodeficient scid mice, subcutaneously inoculated HT-29 xenograft tumors grew more rapidly in obese than lean mice [30]. Reduction of circulating TNF-α levels resulted in suppressed IR signaling in tumor tissue and significantly alleviated tumor growth in ob/ob mice. When ob/ob mice were crossed with transgenic mice overexpressing TGF-α, a ligand for epidermal growth factor receptor, specifically in the mammary epithelium, a significantly reduced incidence of breast tumors was observed in leptin-deficient tumor-bearing mice.

Taken together, these data indicate that while ob/ob mice represent a valuable tool in obesity and diabetes research, it may not be an ideal model for cancer-associated studies since leptin may be required for tumor development. Indeed, recent data demonstrate that leptin signaling is required for the expansion of tumor-initiating cells, at least in the setting of hepatocellular carcinoma.

The leprdb/leprdb (db/db) mice harbor another autosomal recessive mutation in the diabetes (db) gene which was eventually mapped to the leptin receptor (lepr or obr) gene locus. In contrast to leptin-deficient ob/ob mice, db/db animals have high circulating leptin levels resulting from leptin resistance. db/db mice also exhibit obesity, hyperglycemia, and hyperinsulinemia.

Zucker fatty (fa/fa) rats represent a commonly used rat model of genetically induced obesity. fa/fa rats harbor the autosomal recessive fatty (fa) mutation which, by analogy with the db allele, has also been mapped to the obr gene locus located on chromosome 5 in rats. Rats homozygous for the fa mutation display obesity, leptin and insulin resistance, hyperleptinemia and hyperinsulinemia.

LA/N-cp (corpulent) rat is another rodent model of genetically induced obesity that has been used (although less commonly) to study the impact of obesity, diabetes, and insulin resistance on tumor development. A homozygous recessive mutation in the corpulent gene (cp/cp) (also mapped to the obr gene locus) results in excess weight, hyperinsulinemia, and hyperlipidemia.

After intravenous inoculation of mouse melanoma B16 cells, pulmonary metastasis is enhanced in db/db mice. In addition, depletion of NK cells by an anti-asialo-GM1 antibody abrogated the enhanced metastasis in db/db mice.

To understand the impact of obesity on gastrointestinal tumors, chemically and genetically induced colon carcinogenesis was employed in leptin receptor-deficient rodent models. Growth of colorectal and gastric tumors was significantly enhanced in AOM- and NMU-treated db/db mice compared to the respective WT controls.

Food restriction in db/db mice resulted in normalized body weights and blood glucose levels; however, an increased number of premalignant lesions in colon was observed along with hyperinsulinemia and hyperleptinemia suggesting that either of these factors may be responsible for the increased tumor growth.

fa/fa rats are also more susceptible to chemically induced colon carcinogenesis when compared to lean controls, similar to db/db and ob/ob mice [31,32]. A moderate energy restriction that did not affect body weight resulted in the significantly reduced number of advanced crypt foci (ACF) in the colon of fa/fa rats. Similarly, pair feeding of fa/fa rats with lean control animals resulted in the reduced body weight and decreased ACF number in fa/fa rats. These dietary manipulations, however, did
not significantly affect hyperinsulinemia, thus suggesting that insulin may not be a major determinant of the tumor-promoting action of obesity on colon carcinogenesis in this specific model.

In addition, when db/db mice were crossed with APC\(^{1638N/+}\) mice, these animals had a higher number of tumors compared to APC\(^{1638N/+}\) mice. Tumors also had ectopic location (stomach and large intestine) in the double mutants (APC\(^{1638N/+}\)/db/db) but not in the APC\(^{1638N/+}\) mice [33]. APC\(^{1638N/+}\) mice have a mutation in the tumor suppressor APC gene resulting in the enhanced signaling through the Wnt/beta-catenin pathway. This model has a high clinical relevance as the majority (>90%) of patients with colorectal carcinoma harbor somatic mutations in genes encoding individual components of the Wnt/beta-catenin pathway.

The development of 7,12-dimethylbenz(a)anthracene (DMBA)-induced breast tumors is associated with a shorter tumor latency and a more aggressive phenotype in Zucker and corpulent female rats [32]. In contrast, the development of NMU-induced breast cancer is suppressed in female fa//fa rats [34,35]. This discrepancy has been attributed to the different mode of action of the two chemical carcinogens.

In genetically induced models of breast cancer, db/db mice also displayed attenuated tumor progression and/or metastasis when crossed with MMTV-PyMT (polyoma virus middle T antigen) or MMTV-TGF-\(\alpha\) mice [36,37]. It is important to note that breast tumors in these models lack leptin receptors that apparently are critical for tumor development in this setting. This observation has been corroborated by a study with transplantable mammary tumors. When breast tumors induced by overexpression of Wnt transgene were transplanted into db/db mice, their growth was markedly accelerated in the setting of hyperleptinemia, whereas transplantation into leptin-deficient ob/ob mice resulted in the delayed tumor growth [38].

Taken together, the data obtained in the aforementioned models clearly indicate that obesity, diabetes and/or insulin resistance promote gastrointestinal carcinogenesis, although specific pathophysiological mechanisms mediating this effect have not been clearly dissected. Studies aimed to elucidate the effect of obesity and diabetes on breast cancer and melanoma in leptin and/or leptin-receptor-deficient models should be interpreted with caution. A critical role of leptin signaling in normal and malignant mammary cells makes ob and obr-deficient models inappropriate to study breast carcinogenesis.

**Diet-induced obesity (DIO) and cancer**

Different types of high energy/high fat diet (HED or HFD), in which up to 60% calories are coming from fat, have been extensively employed in obesity and diabetes research. Animal models utilizing HFD and different tumor models are particularly helpful in understanding a relationship between obesity, insulin resistance, and tumor development. The major limitation of DIO models, similar to models of genetically induced obesity, is that in most cases they are unable to dissect a specific pathophysiologic mechanism mediating obesity-mediated tumor progression, mostly in the setting of breast, prostate, and colon.

In most experimental models of breast cancer including NMU- and DMBA-induced models of chemically induced carcinogenesis as well as transgenic and implantation breast cancer models, DIO results in enhanced tumor development and progression. HFD promotes NMU-induced mammary carcinogenesis as demonstrated by shortened latency and increased multiplicity of tumors in Sprague-Dawley and Fischer 344 (F344) rats [39], whereas caloric restriction, physical exercise and dietary supplementation with fibers inhibit the stimulatory effect of HFD on NMU-induced tumor development [40,41]. DMBA-induced breast carcinogenesis is also enhanced in Sprague-Dawley rats fed a HFD [42], and ovariectomy further enhances this effect [43]. In C3H mice, obesity induced by a single injection of gold thioglucose, which induces necrosis in the ventromedial portion of hypothalamus, results in accelerated development of spontaneous mammary tumors, and HFD further enhances the incidence of mammary tumors in this model. Furthermore, while ovariectomized lean C3H female mice do not develop breast tumors, ovariectomized obese mice exhibit an increased tumor incidence compared to non-ovariectomized obese controls. Similar data were reported in the syngeneic Wnt-1 model of breast cancer. Ovariectomy inhibits growth of orthotopically inoculated estrogen-sensitive Wnt-1 cells in C57Bl/6 mice fed regular chow, whereas ovariectomized mice fed HFD further gain weight and display accelerated tumor growth, which is abrogated by calorie restriction and pharmacologic mTOR inhibition [44,45]. These data suggest that ovarian hormones have a different impact on tumor development in lean and obese animals. Obesity enhances tumors growth in the absence of ovarian hormones and this affect is mediated by the PI3K/Akt/mTOR pathway.

The stimulatory effect of DIO on tumor progression has been demonstrated in several transgenic mouse models of breast cancer. In the transgenic MMTV-v-Ha-ras and MMTV-TGF-\(\alpha\) models, HFD increases the incidence and shortens the latency of breast tumors. HFD also significantly enhances growth of primary mammary tumors and pulmonary metastases when MMTV-driven polyoma virus middle T antigen (PyVnT) was introduced into M16i mice, which represent a model of polygenic obesity.

The effect of HFD on HER2/ErbB2-induced mammary tumorigenesis is not completely understood. HFD does not affect tumor latency, incidence, burden and metastasis in mice over-expressing wild-type Neu (mouse analogue of human HER2) in mammary epithelium but enhances tumor growth in MMTV-c-Neu mice that overexpress a constitutively active form of Neu [46,47].

In a xenograft LNCaP model, growth of subcutaneously inoculated tumor cells was significantly enhanced in nude mice fed a high calorie diet. This effect is accompanied by elevated insulin and IGF-I levels in the circulation, increased insulin receptor (IR) and IGF-I expression and enhanced Akt phosphorylation...
in the tumor tissue [48]. However, tumor incidence is decreased in transgenic Hi-Myc and xenograft LNCaP models of prostate cancer when these mice are fed a low-fat diet. These animals display elevated IGFBP-1 levels in the circulation and reduced Akt phosphorylation in the tumor tissue, indicative of suppressed IR signaling [49,50]. Taken together, these data suggest a potential role of hyperinsulinemia and enhanced IR/IGF-IR signaling in the stimulatory effects of obesity on prostate tumor growth. In a model of AOM-induced colon carcinogenesis, HFD enhances ACF formation in F344 rats. HFD also increased the number of AOM-induced colon polyps in adiponectin and adiponectin receptor knockout mice compared to wild-type controls [51]. Administration of exogenous adiponectin inhibits growth of implanted tumors in mice fed a HFD, possibly through a direct regulation of cell proliferation, adhesion, invasion and anchorage-independent growth and modulation of several key tumorigenic signaling pathways (AMPK/mTOR, STAT3, VEGF, p21/p53/cyclin) [52]. However, transgenic overexpression of adiponectin fails to delay AOM-induced colon carcinogenesis. The results of these studies indicate that specific stimulatory effects of obesity on tumor growth can be mediated, at least in part, by hypoadiponectinemia, one of the hallmarks of obesity.

In the APC<sup>1638G+/-</sup> mouse model of intestinal tumorigenesis, a Western-style diet containing high fat and phosphate but low calcium and vitamin D, enhances intestinal tumorigenesis. This diet also triggers adenoma formation in wild-type mice and significantly increases incidence, multiplicity and size of small and large intestinal adenomas in p27<sup>−/−</sup>-null mice. An exposure of p27<sup>−/−</sup>-null mice to HFD also results in the development of malignant tumors.

In the syngeneic colon carcinoma MC38 model, C57Bl/6 mice fed a HFD demonstrate increased tumor growth and metastasis. This effect is abolished in mice with genetically disrupted igf1 gene in the liver resulting in chronic IGF-I deficiency [53]. Furthermore, the growth of implanted MC38 tumors in obese mice is enhanced by ovariectomy [54]. In ovariectomized animals fed a HFD, tumor growth was associated with insulin and leptin resistance as well as higher levels of pro-inflammatory proteins. In line with the aforementioned data on breast cancer, this study thus strongly supports the hypothesis that ovarian hormones have a modulatory effect on obesity-induced insulin resistance and inflammation, which may directly or indirectly influence colorectal tumor growth.

HFD also enhances growth of Lewis lung carcinoma (LLC) cells in C57Bl/6 mice [54]. In addition to HFD, high sucrose (HSD) and high cholesterol diets (HCD) are also capable of stimulating LLC growth and metastasis in recipient C57Bl/6 mice, but only HFD results in the development of obesity, insulin resistance, and hyperinsulinemia suggesting that factors other than insulin mediate tumor-stimulating effects of HSD and HCD.

HFD triggers the development of spontaneous hepatomas and enhances diethyltinlsosamine-induced hepatocarcinogenesis in C3H mice and Sprague-Dawley rats, respectively. A diet enriched in fat increases the incidence of DMBA-induced pancreatic cancer in Sprague-Dawley rats. In “Rockland strain” mice, HFD promotes benzpyrene-induced skin tumorogenesis. Rapid body weight gain also increases the risk of UV radiation-induced skin carcinogenesis in SKH-1 hairless mice.

Taken together, these data indicate a stimulatory effect of DIO on tumor growth. However, most of the studies described earlier do not dissect a specific pathophysiologic mechanism mediating this effect although several lines of evidence indicate that IR/IGF-IR and/or adiponectin signaling may play a role in this process.

**Non-obese models of type 2 diabetes and cancer**

A-ZIP/F-1 or fatless mouse represents a transgenic mouse model of lipoatrophic diabetes which was employed to dissect effects of body adiposity and diabetes on tumor growth. Skin and breast carcinogenesis is remarkably enhanced in A-ZIP/F-1 mice compared to the respective controls. These findings indicate that in the absence of fat tissue diabetes per se has a direct stimulatory effect on malignant transformation and tumor growth. The results of this study, however, should be interpreted with caution because (i) lipoatrophic form of T2DM is rare in humans, (ii) A-ZIP/F-1 mice have severely impaired mammary gland development, and (iii) the spectrum of metabolic abnormalities in A-ZIP/F-1 mice is very diverse and includes hyperinsulinemia, severe hyperglycemia, hyperlipidemia, elevated levels of circulating IGF-I and proinflammatory cytokines, and each of these factors may be capable of enhancing tumor growth. In addition, tumor tissue in A-ZIP mice has increased levels of phosphorylated ERK1/2 and Akt, whereas IR phosphorylation is attenuated suggesting that other factors rather than the IR are responsible for enhanced signaling though the PI3K and MAPK pathways in this model.

MKR mouse represents a unique transgenic model of T2DM. In this model, a mutant IGF-IR (IGF-IR<sup>K1003R</sup>) is exclusively overexpressed in the skeletal muscle under the control of a muscle creatine kinase promoter. The overexpression results in the abrogation of insulin and IGF-I/II signaling due to the formation of hybrids between dominant-negative IGF-IR<sup>K1003R</sup> and endogenous insulin and IGF-I receptors. MKR females display a mild diabetic phenotype; they develop severe insulin resistance and hyperinsulinemia in the setting of moderately reduced body adiposity and mild dysglycemia and hyperlipidemia [55]. This model is particularly useful in understanding the effects of hyperinsulinemia and insulin resistance on tumor growth. MKR female mice demonstrate a more aggressive phenotype and enhanced tumor growth in transgenic MMTV-PyVmt and syngeneic Met-1 and MCNeuA models of breast cancer [55]. In addition, MKR mice display markedly enhanced pulmonary metastasis after orthotopic and intravenous inoculation of Mvt-1 cells and in transgenic doxycycline-inducible MMTV-rtTA/tetO-NeuNT model of breast cancer [56]. Tumor tissue extracted from MKR mice displays enhanced phosphorylation of the IR/IGF-IR, Akt,
S6 ribosomal protein and upregulation of c-myc expression suggesting a direct effect of hyperinsulinemia on the IR and/or the IGF-IR with subsequent activation of the PI3K/Akt/mTOR pathway and c-myc-driven expression in tumor cells [55–57]. Furthermore, pharmacologic correction of insulin resistance and hyperinsulinemia by insulin-sensitizing agent and pharmacologic blockade of insulin and IGF-I receptors by BMS-536924 abrogate tumor-promoting effects of hyperinsulinemia in MKR mice [55]. In addition, administration of the PI3K/Akt/mTOR pathway by rapamycin, NVP-BEZ235 and NVP-BKM-120, small molecule inhibitors which block the pathway at different levels, is also capable of preventing accelerated tumor growth in MKR mice [57,58]. Collectively, these data corroborate a key pathophysiologic role that insulin and its cognate receptor play in the setting of insulin resistance and identify downstream signaling molecules that mediate this effect.

Conclusions and future directions

While there is compelling epidemiologic evidence of the association of diabetes and cancer risk and experimental evidence is being obtained regarding the potential factors that may be causative, much more research on this topic is to be encouraged. More studies are required to show the relationship across causative, much more research on this topic is to be encouraged. Finally, both animal and human studies will be necessary to fully determine which cancers are involved. More studies are required to show the relationship across causative, much more research on this topic is to be encouraged. Furthermore, pharmacologic correction of insulin resistance and hyperinsulinemia by insulin and its cognate receptor play in the setting of insulin resistance and identify downstream signaling molecules that mediate this effect.

References


Diabetes and sleep apnea

Abd A. Tahrani
Centre of Endocrinology, Diabetes and Metabolism (CEDAM), School of Clinical and Experimental Medicine, University of Birmingham, Birmingham, UK

Key points
- Obstructive sleep apnea (OSA) is common and associated with increased risk of dysglycemia, type 2 diabetes, hypertension, cardiovascular disease, and mortality.
- Patients with type 2 diabetes and insulin resistance are at increased risk of developing OSA.
- OSA is associated with activation of the hypothalamic pituitary adrenal axis, adipokine changes, autonomic dysfunction, increased inflammation and oxidative stress, and endothelial dysfunction which might explain the association between OSA and dysglycemia.
- CPAP (continuous positive airway pressure) treatment improves many of the molecular and hormonal consequences of OSA and has a beneficial impact on OSA comorbidities, but the impact of CPAP in patients with type 2 diabetes remains unclear.

Introduction

Obesity is a global epidemic that has far-reaching economic and public health implications [1]. Excess body weight is associated with multiple comorbidities including type 2 diabetes (T2DM), hypertension, obstructive sleep apnea (OSA), and cardiovascular disease (CVD) [2]. Dysglycemia and OSA share common risk factors such as age and obesity; hence it is not surprising that OSA and glycemic abnormalities often coexist.

Despite the close association between OSA and dysglycemia, the direction of this association and whether obesity modulates the association remain unclear. The interaction between multiple genetic and environmental factors gives rise to a heterogeneous and progressive condition with variable degrees of insulin resistance (IR) and pancreatic β-cell dysfunction [3]. When β cells are no longer able to secrete sufficient insulin to overcome IR, impaired glucose tolerance (IGT) progresses to T2DM [3]. OSA is associated with different factors that alter IR and β-cell function suggesting that OSA can result in glycemic abnormalities. On the other hand, dysglycemia and its association with inflammation and autonomic dysfunction can lead to or worsen pre-existing sleep apnea [4].

An increasing number of human and animal studies have expanded our understanding of the association between OSA and dysglycemia and how each condition modulates the metabolic impacts of the other [5]. In this chapter we aim to review the evidence for the association between OSA and dysglycemia, to explore the underlying mechanisms and to assess the metabolic impacts of OSA in the presence of dysglycemia and vice versa. The mainstay of this chapter will focus on OSA, rather than central sleep apnea (CSA), as it is the most common type of sleep apnea and the one that is closely related to obesity and T2DM. We will, however, highlight some important issues in relation to CSA when needed.

Definitions

Sleep apnea can be either central or obstructive. OSA is a common medical disorder that is characterized by instability of the upper airway during sleep, which results in markedly reduced (hypopnea) or absent (apnea) airflow at the nose or mouth [6]. These apnea/hypopnea episodes are usually accompanied with oxygen desaturations and micro arousals that cause sleep fragmentation and reduction in slow wave and rapid eye movement (REM) sleep [6]. The American Academy of Sleep Medicine (AASM) 2007 guideline has defined sleep events apnea as cessation or ≥90% reduction in airflow for a period of ≥10 seconds and hypopnea as ≥30% reduction in airflow for ≥10 seconds associated with ≥4% drop in oxygen saturations; an alternative definition of hypopnea based on ≥50% reduction in airflow with ≥3% drop in oxygen saturations can also be used [7]. The updated AASM 2012 guidelines defined hypopnea in adults as ≥30% drop of air flow of pre-event baseline for ≥10 seconds in association with either ≥3% arterial oxygen desaturation or an arousal [8]. Apneas are classified into obstructive or central...
Diabetes and sleep apnea

Figure 22.1 An example of apnea (red), hypopneas (pink), and desaturations (green) in a patient with OSA and T2DM. Note the presence of thoracic and abdominal movements indicating the presence of obstructive rather than central sleep apnea. (For a color version of this figure, please see the color plate section.)

based on the presence or absence of respiratory/abdominal efforts. An example of apneas and hypopneas can be found in Figure 22.1.

The apnea-hypopnea index (AHI) is the average number of apneas and hypopneas episodes per hour during sleep and is a marker of the severity of OSA [6]. An AHI of $\geq 5$ events per hour is consistent with the diagnosis of OSA [9]. OSA can be classified to mild, moderate, and severe based on AHI 5 to $<15$, 15 to $<30$, and $\geq 30$ events per hour. The respiratory disturbance index (RDI) is another OSA measure that includes the AHI in addition to respiratory effort-related arousal, which is defined as a sequence of breaths characterized by increasing respiratory effort leading to an arousal from sleep, but which does not meet criteria for an apnea or hypopnea [6]. Another measure of OSA is the oxygen desaturation index (ODI), which is the average number of oxygen desaturations per hour during sleep.

**OSA epidemiology and risk factors**

The prevalence of OSA varies considerably between studies, mainly due to differences in the population studied, study designs, and the method and criteria used to diagnose OSA. A prevalence of 4% in men and 2% in women [6] has traditionally been quoted in many populations. The prevalence from three well-conducted studies with similar design from Wisconsin, Pennsylvania, and Spain showed an OSA prevalence of 17–26% in men and 9–28% in women and a prevalence of 9–14% and 2–7% for men and women with moderate to severe OSA [10]. These studies used a two-stage sampling design which allows some degree of estimate of the “self-selection” bias which is usually a significant problem in OSA studies.

OSA prevalence is affected by many risk factors such as obesity, age, ethnicity, and gender. The large majority of epidemiologic OSA studies took place in Western societies (particularly in the USA) and included mainly white Caucasians; hence data about the prevalence in other ethnicities is rather limited. In African Americans, the available results are conflicting. While some studies showed a higher adjusted OSA prevalence in Afro-Caribbeans (increased twofold) [10], others did not show such a difference [11]. Chinese have a high OSA prevalence (8.8% in men and 3.7% in women) despite being less obese than white Europeans [12,13]. This highlights the importance of factors other than obesity (such as the anatomy of upper airways) in the development of OSA. Chinese were shown to have more crowded upper airways with higher Mallampati score and shorter thyromental distance [14]. Data regarding the prevalence of OSA in South Asians are also limited. In a semi-urban population in Delhi, the OSA prevalence was 3.7% [15], this rose to 9.3% in middle-aged urban Indians [16], and 19.5% in middle-aged urban men [17].

The impact of gender on OSA status has been well recognized; men have two to three times increased risk of OSA compared to women [10]. The exact mechanisms behind the
gender differences in OSA prevalence are not clear but several possible factors have been proposed. Sex hormones have been implicated, since men receiving testosterone replacement are at higher risk of OSA and the prevalence of OSA in postmenopausal women is higher than in premenopausal women. Hormone-replacement therapy also reduces the risk of OSA in postmenopausal women [18]. In addition, hyperandrogenemia was associated with higher risk of OSA and oral contraceptives were associated with lower OSA risk in women with polycystic ovarian syndrome [19]. Differences in upper airway size and ventilator control between men and women have also been implicated but the results are conflicting [20].

Several studies have shown that the prevalence of OSA increases with age [18]. In men, OSA (AHI ≥10 events per hour) was present in 3.2%, 11.3%, and 18.1% of the 20–44, 45–64, and 61–100-year-old age groups, respectively [21]. In another study from Spain, the prevalence of any OSA was three times higher and the prevalence of moderate to severe OSA was four times higher in older patients (>70 years old) compared to middle-aged participants [10]. On the other hand, in the Sleep Heart Health Study, OSA prevalence increased with age but reached a plateau at the age of 65 years, which may be related to survival effect [11]. The relationship with age seems to be due to changes in pharyngeal anatomy and upper airway collapsibility [18].

Excess body weight is an essential risk factor for OSA, although not all OSA patients are obese or overweight. In the Wisconsin Sleep Cohort Study, each increase in BMI by one standard deviation, resulted in a fourfold increase in OSA prevalence [22]. Several other studies have shown the strong link between OSA and excess body weight [18]. Prospective studies showed that weight gain is associated with the development of or worsening pre-existing OSA [23,24]. This was further supported by a randomized controlled trial which showed that weight loss (vital life style modifications or surgical intervention) improve/cure OSA [25,26]. The mechanisms that link obesity to OSA are not entirely clear but several mechanisms have been proposed; weight gain can alter normal upper airway mechanics during sleep by increased parapharyngeal fat deposition resulting in a smaller upper airway, altering the neural compensatory mechanisms that maintain airway patency, reducing the functional residual capacity with a resultant decrease in the stabilizing caudal traction on the upper airway and affecting the chemosensitivity to O₂ and CO₂ which reduces ventilator drive [27].

There are several other predisposing risk factors to OSA such as current smoking, excess alcohol intake, and genetic factors [10,18].

### OSA pathophysiology

OSA is a very complex disorder, and although obesity and fat deposition around the neck plays an important role, there are many other important players that contribute to the development of this condition.

The human upper airway is a unique multipurpose structure involved in performing a variety of tasks such as speech, swallowing, and the passage of air for breathing [28]. The airway, therefore, is composed of numerous muscles and soft tissue but lacks rigid or bony support [28]. Most notably, it contains a collapsible portion that extends from the hard palate to the larynx which allows the upper airway to change shape and momentarily for speech and swallowing during wakefulness; but this feature also provides the opportunity for collapse at inopportune times such as during sleep [28]. Several imaging-based studies showed that patients with OSA have a smaller upper airway during wakefulness and anesthesia, resulting in an airway that is more prone to collapse [28]. The upper airway muscles (genioglossus) activity is also increased in OSA patients compared to age- and obesity-matched healthy controls [29], suggesting that these muscles are compensating for an underlying defect in the anatomy of the upper airway in patients with OSA [28]. This muscle hyperactivity is resolved in CPAP-treated patients [30]. Sleep onset is associated with greater reductions in upper airway muscles tone in OSA patients, which explains the occurrence of apnea/hypopnea episodes at sleep onset and during REM sleep [28]. This reduction in upper airway muscle tone during sleep seems to result from a central lack of drive and local inhibitory reflexes that respond to changes in pressure in the upper airways [28]. Several studies have also shown that changes in lung volume affect upper airway muscles activity [28]. Other abnormalities described in OSA include irregularities in ventilatory control and stability, changes in chemosensitivity to CO₂, and higher arousal thresholds [28].

### OSA clinical features and diagnosis

Good history and examination are still an essential part of the assessment of patients with OSA despite the fact that several reports have shown the limited value of symptoms in predicting OSA (in one report, only one third of patients would have been identified clinically) [31]. Snoring is the most common symptom of OSA and it occurs in 95% of patients [6]. Snoring, however, has a poor predictive value due to the high prevalence of snoring and the fact that many snorers do not have OSA [6]. Nonetheless, lack of snoring almost rules out OSA since only 6% of OSA patients (or their partners) have not reported snoring [6]. Witnessed apneas are another important symptom that is usually reported by the partner. However, witnessed apneas do not correlate with disease severity and up to 6% of the “normal” population could have witnessed apneas without OSA [6]. Other nocturnal symptoms such as choking (which is possibly a “proper” rather than a “micro” arousal to terminate apnea), insomnia, nocturia, and diaphoresis have been reported [6]. Daytime symptoms include excessive...
daytime sleepiness, fatigue, morning headache, and autonomic symptoms [6].

The gold standard to diagnose OSA is polysomnography that typically includes the recording of 12 channels such as EEG, electro-oculogram (EOG), electromyogram (EMG), oronasal airflow, chest wall effort, abdominal effort, body position, snore microphone, ECG, and oxyhemoglobin saturation [6]. The main problem with polysomnography is that it is time consuming and expensive. Portable home-based respiratory devices are another alternative [6]. The main advantages are that they are less resourceful but they are associated with higher failure/loss of lead rate compared to polysomnography [6]. Pulse oximetry is another good way to diagnose OSA, it cannot, however, differentiate between obstructive and central apneas and it has a wide range of sensitivity (31–98%) and specificity (41–100%). The AASM recommends use of a Type III device as a minimum [6].

**OSA and glucose metabolism**

OSA and dysglycemia have similar risk factors (namely obesity) and hence it is not surprising that these conditions co-exist. However, not all obese patients have both conditions and many patients have one and not the other. Hence, understanding this association and the mechanisms that underpin this relationship is important to understand the pathogenesis of OSA and T2DM. There are many studies that have examined the association between snoring, as a surrogate marker of OSA, and different aspects of glucose metabolism [32]; here, however, we will mainly focus on studies that validated the presence and severity of OSA using more accurate methods.

**OSA and prediabetes**

OSA has been associated with components of the metabolic syndrome and with IR independent of obesity [33]. Some studies reported that abnormal glycemia could occur in as many as 79% of patients with OSA [34]. In a cross-sectional analysis in a subset of the Sleep Heart Health Study, relative to those with RDI < 5, individuals with mild and moderate to severe OSA had adjusted OR of 1.27 (95% CI 0.98 – 1.64) and 1.46 (95% CI 1.09 – 1.97), respectively for fasting glucose intolerance [35]. Sleep-related hypoxemia was also associated with glucose intolerance independently of age, gender, BMI, and waist circumference [35]. In another study that included 150 men not known to have diabetes, an AH1 ≥ 5 was associated with an increased risk of IGT or DM (based on OGTT) (OR: 2.15; 95% CI 1.05 – 4.38) following adjustment for BMI and body fat (as measured by hydrodensitometry) [36]. For each 4% decrease in oxygen saturation, the OR of worsening glucose tolerance was 1.99 (95% CI 1.11 – 3.56) after adjusting for percent body fat, BMI, and AH1 [36]. Similarly, in a cross-sectional analysis of 2588 participants (aged 52–96 years; 46% men), the OSA group (RDI ≥ 10) had higher adjusted OR of 1.3 (95% CI 1.1 – 1.6) for IFG (impaired fasting glucose), 1.2 (1.0 – 1.4) for IGT, 1.4 (1.1 – 2.7) for IFG plus IGT, and 1.7 (1.1 – 2.7) for occult diabetes compared to those without OSA [37]. In overweight/obese individuals, FPG and the 2-h OGTT glucose values were significantly higher in patients with vs. without OSA while there was no such difference in normal-weight subjects [37]. In the same study, the prevalence of abnormal glycemia was significantly higher in normal (679 subjects) and overweight/obese (1909 subjects) individuals with OSA compared to those without (no OSA vs. OSA: 17.6% vs. 25.5%, p = 0.03 and 36.9% vs. 43%, p = 0.02 for IFG; 5.1% vs. 9.3%, p = 0.06 and 13.6% vs. 17.3%, p = 0.05 for combined IFG and IGT; 5.7% vs. 9.3%, p = 0.09 and 9.3% vs. 13.2%, p = 0.01 for normal and overweight/obese subjects, respectively) [37]. More recently, the higher AH1 and lower nocturnal hypoxemia were shown to be associated with higher HbA1c in patients without diabetes despite adjustment for a wide range of confounders in a cross-sectional study [38].

**OSA and insulin resistance**

Several studies examined the association between OSA and IR (Table 22.1). The majority of these studies were conducted in white Caucasians, but some included patients of other racial groups and ethnicities. The major difficulty in examining the association between OSA and IR is teasing out the impact of OSA from that of obesity. Most studies showed an association between OSA and IR, but some did not (Table 22.1). Studies that did not show such a relationship included fewer participants and potentially were underpowered.

Interestingly, a small recent study suggests that excessive daytime sleepiness may be, in part, responsible for the IR in patients with OSA [39]. Barcelo et al. studied 44 patients with OSA (22 with and 22 without excessive daytime sleepiness) matched for age, BMI and AH1, and 23 healthy controls [39]. Patients with excessive daytime sleepiness (assessed by the Epworth Sleepiness Scale and the Multiple Sleep Latency Test) had higher HOMA-IR compared with OSA patients without excessive daytime sleepiness or healthy controls [39]. The difference in HOMA-IR between OSA patients without excessive daytime sleepiness and healthy controls was not significant [39]. Glucose levels were significantly higher in patients with OSA and excessive daytime sleepiness compared to those with OSA without excessive daytime sleepiness and healthy controls [39]. In support for the association between excessive daytime sleepiness and IR, CPAP treatment in the same study reduced the HOMA-IR and increased IGF-1 levels in patients with excessive daytime sleepiness, but did not modify any of these variables in patients without excessive daytime sleepiness [39]. This impact of excessive daytime sleepiness may explain in part some of the variation in the association between OSA and IR observed in cross-sectional studies. Recent studies tried to exclude the impact of obesity on IR by examining healthy lean men and found that OSA was associated with IR despite the lack of obesity (Table 22.1) [40, 41].
<table>
<thead>
<tr>
<th>Study/ year</th>
<th>Study population</th>
<th>Sample size</th>
<th>OSA assessment</th>
<th>IR assessment</th>
<th>Confounders assessed</th>
<th>Results</th>
<th>Association</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tiihonen et al. [43] 1993</td>
<td>Suspected OSA</td>
<td>18</td>
<td>Oximetry</td>
<td>IR index during OGTT</td>
<td>Age, BMI</td>
<td>IR independently associated with ODI after adjustment</td>
<td>+</td>
</tr>
<tr>
<td>Strohl et al. [44] 1994</td>
<td>Suspected OSA, all men, mean age 47 years</td>
<td>261</td>
<td>PSG</td>
<td>Fasting insulin</td>
<td>BMI</td>
<td>IR correlated with AHI independent of BMI</td>
<td>+</td>
</tr>
<tr>
<td>Davies et al. [45] 1994</td>
<td>Confirmed OSA and control for snorers</td>
<td>15 OSA/ 18 control</td>
<td>PSG</td>
<td>Fasting insulin</td>
<td>Matched for age, sex, BMI, smoking, and alcohol</td>
<td>No difference in IR between groups</td>
<td>–</td>
</tr>
<tr>
<td>Grunstein et al. [46] 1995</td>
<td>The Swedish Obese Subjects cohort</td>
<td>3034</td>
<td>Questionnaire</td>
<td>Fasting insulin</td>
<td>BMI, visceral fat, age</td>
<td>IR higher in patients with high risk for OSA after adjustment</td>
<td>+</td>
</tr>
<tr>
<td>Stroohs et al. [47] 1996</td>
<td>Healthy individuals (34 women) with a mean age of 44.3 years and a mean BMI of 27.1</td>
<td>50</td>
<td>Validated sleep apnea recorder</td>
<td>Insulin suppression test with somatostatin analogue</td>
<td>BMI</td>
<td>IR was higher in individuals with ODI ≥ 10 h⁻¹ but the difference disappeared after adjusting for BMI</td>
<td>–</td>
</tr>
<tr>
<td>Saarelainen et al. [48] 1997</td>
<td>Randomly selected OSA patients from sleep clinic, no controls</td>
<td>10</td>
<td>PSG</td>
<td>Euglycemic hyperinsulinemic clamp</td>
<td>Age, BMI</td>
<td>No association between IR and OSA</td>
<td>–</td>
</tr>
<tr>
<td>Ip et al. [49] 2000</td>
<td>OSA patients and controls from a sleep clinic in Hong Kong</td>
<td>30/30</td>
<td>PSG</td>
<td>Fasting insulin</td>
<td>Matched for BMI, age, sex, menopausal status</td>
<td>OSA correlated with IR after adjustment</td>
<td>+</td>
</tr>
<tr>
<td>Vgontzas et al. [50] 2000</td>
<td>Obese men with OSA and control groups</td>
<td>37</td>
<td>PSG</td>
<td>Fasting insulin</td>
<td>OSA were age and BMI matched to nonapneic individuals and age-matched lean men</td>
<td>OSA patients were more insulin resistant compared to nonapneic patients. However, visceral fat was higher in patients with OSA despite similar BMI to those without OSA</td>
<td>+</td>
</tr>
<tr>
<td>Elmasry et al. [51] 2001</td>
<td>Hypertensive men</td>
<td>116</td>
<td>PSG</td>
<td>Fasting insulin</td>
<td>Waist:hip ratio, age</td>
<td>OSA was associated with IR after adjustment</td>
<td>+</td>
</tr>
<tr>
<td>Ip et al. [52] 2001</td>
<td>Consecutive Chinese subjects (197 men) from the Hong Kong sleep clinic</td>
<td>270</td>
<td>PSG</td>
<td>HOMA-IR</td>
<td>BMI, age</td>
<td>AHI and minimum oxygen saturation independently associated IR; this association was seen in obese and nonobese individuals. Each additional AHI unit per sleep hour increased the fasting insulin level and HOMA-IR by about 0.5%</td>
<td>+</td>
</tr>
<tr>
<td>Study</td>
<td>Population/Study Details</td>
<td>Methods</td>
<td>Main Findings</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------------</td>
<td>------------------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manzella et al. [53] 2002</td>
<td>OSA patients, mean BMI 30.9 kg/m²</td>
<td>20 Validated sleep recorder, Euglycemic hyperinsulinemic clamp</td>
<td>AHI correlated with insulin-mediated glucose uptake ($r = -0.73, p &lt; 0.001$), remained significant after adjustment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Punjabi et al. [36] 2002</td>
<td>Community sample, all men, mean BMI 30.5 kg/m²</td>
<td>151 PSG HOMA-IR</td>
<td>Increasing AHI was associated with worsening insulin resistance independent of obesity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meslier et al. [54] 2003</td>
<td>All men suspected with OSA</td>
<td>595 PSG HOMA-IR</td>
<td>Insulin sensitivity decreases as OSA severity increased</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tassone et al. [55] 2003</td>
<td>Obese OSA patients, mean BMI 38.6, mean AHI 40.5 compared to weight-matched nonapneic controls</td>
<td>30/27 PSG ISI</td>
<td>ISI was lower in OSA patients</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barcelo et al. [56] 2004</td>
<td>OSA and control</td>
<td>65 PSG HOMA-IR</td>
<td>No association</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Punjabi et al. [35] 2004</td>
<td>Community sample</td>
<td>2656 PSG HOMA-IR</td>
<td>Moderate to severe OSA was associated with IR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Makino et al. [57] 2006</td>
<td>Japanese patients with OSA</td>
<td>213 PSG HOMA-IR</td>
<td>AHI independently associated with IR after adjustment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gruber et al. [58] 2006</td>
<td>Suspected OSA</td>
<td>79 PSG HOMA-IR</td>
<td>OSA was associated with IR but not after adjustment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McArdle et al. [59] 2007</td>
<td>Moderate to severe OSA patients with controls</td>
<td>42 PSG HOMA-IR</td>
<td>OSA associated with higher HOMA-IR, remained significant after adjustment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sharma et al. [60] 2007</td>
<td>OSA patients with obese and nonobese controls</td>
<td>40/40/40 PSG HOMA-IR</td>
<td>No difference in IR across groups</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Onat et al. [61] 2007</td>
<td>Turkish population</td>
<td>1946 symptoms HOMA-IR</td>
<td>No association</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kapsimalis et al. [62] 2008</td>
<td>Suspected OSA</td>
<td>67 PSG HOMA-IR</td>
<td>No association</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Theorell-Haglow et al. [63] 2008</td>
<td>Population-based, all women</td>
<td>400 PSG ISI</td>
<td>Nocturnal minimal saturation was independently associated with decreased TB after adjustment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(continual overleaf)
<table>
<thead>
<tr>
<th>Study/ year</th>
<th>Study population</th>
<th>Sample size</th>
<th>OSA assessment</th>
<th>IR assessment</th>
<th>Confounders assessed</th>
<th>Results</th>
<th>Association</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tikacova et al. [64] 2008</td>
<td>Suspected OSA</td>
<td>98</td>
<td>PSG</td>
<td>HOMA-IR</td>
<td>Age, BMI, sex</td>
<td>OSA associated with IR</td>
<td>+</td>
</tr>
<tr>
<td>Punjabi &amp; Beamer [65] 2009</td>
<td>Suspected OSA</td>
<td>118</td>
<td>PSG</td>
<td>Frequent sampling Intravenous glucose tolerance test</td>
<td>Age, sex, race, BMI, body fat</td>
<td>OSA associated with IR. IR increases in a stepwise manner across OSA severity categories</td>
<td>+</td>
</tr>
<tr>
<td>Polotsky et al. [66] 2009</td>
<td>Consecutive individuals with BMI &gt; 40</td>
<td>90</td>
<td>PSG</td>
<td>HOMA-IR</td>
<td>BMI, waist circumference</td>
<td>Oxygen desaturation greater than 4.6% was associated with a 1.5-fold increase in IR</td>
<td>+</td>
</tr>
<tr>
<td>Bhushan et al. [67] 2010</td>
<td>Obese South Asians with OSA and matched control</td>
<td>121/119</td>
<td>PSG</td>
<td>Fasting insulin</td>
<td>Matched for age, BMI, and % body fat</td>
<td>Fasting insulin was associated with OSA</td>
<td>+</td>
</tr>
<tr>
<td>Togeiro et al. [68] 2012</td>
<td>Community sample from Brazil</td>
<td>1042</td>
<td>PSG</td>
<td>HOMA-IR</td>
<td>Age, sex, abdominal obesity, total sleep time</td>
<td>Moderate to severe OSA and time of oxygen saturation &lt;90% were independently associated with impaired fasting glucose, elevated triglycerides, and HOMA-IR</td>
<td>+</td>
</tr>
<tr>
<td>Pamidi et al. [40] 2012</td>
<td>Healthy men, aged 23 years and BMI 22.6 kg m$^{-2}$</td>
<td>52</td>
<td>PSG</td>
<td>Matsuda index and insulin AUC during OGTT</td>
<td>Groups matched for age, BMI, family history of diabetes, exercise</td>
<td>Patients with OSA had lower insulin sensitivity and higher insulin secretion than the control subjects, despite comparable glucose levels</td>
<td>+</td>
</tr>
<tr>
<td>Lin et al. [41] 2012</td>
<td>Individuals with and without OSA, aged 47.5 years, BMI 22.7</td>
<td>113/45</td>
<td>PSG</td>
<td>HOMA-IR</td>
<td>OSA and controls were matched for age, BMI, gender and waist circumference</td>
<td>AHI correlated positively with HOMA-IR</td>
<td>+</td>
</tr>
</tbody>
</table>

ISI, insulin sensitivity index; PSG, polysomnography.
Unlike previous cross-sectional studies, OSA, AHI, ODI, and minimal oxygen saturations were independently associated with IR development over an 11-year follow-up period after adjustment for age, baseline BMI, hypertension, BMI change over follow-up, and CPAP treatment [42].

In the light of the association between OSA and IR, there was a wide interest in the impact of CPAP treatment on IR. Unfortunately the impact of CPAP on IR is not clear. While some studies showed that CPAP treatment lowered IR [69,70], others did not [45,49,71–73]. CPAP treatment varied in these studies from 1 night to 3 years. The lack of a positive effect of CPAP on IR may be the result of several methodological limitations such as the lack of a control group, short duration of CPAP treatment, and small sample sizes. However, it is possible that the impact of long-standing OSA on IR is irreversible. Two recent meta-analyses showed that CPAP treatment was associated with a reduction in HOMA-IR [74,75], although this benefit may occur only in those using CPAP >4h per night [76]. An in-depth review on the impact of CPAP on glucose metabolism can be found in [77].

**OSA and β-cell function**

While the impact of OSA in IR has been studied extensively, the impact of OSA on β-cell function has received little attention, despite being an essential part of the pathogenesis of T2DM and prediabetes. A small number of animal studies showed that intermittent hypoxia increases β-cell death, and results in β-cell dysfunction, although the intermittent hypoxia used in this study is far greater than that which occurs in humans with OSA [78]. One study in humans examined 118 patients without diabetes using the modified frequent sampling intravenous glucose tolerance test and found that the disposition index, a measure of β-cell function, was also reduced in patients with moderate to severe sleep-disordered breathing [65]. Similarly, a more recent study in patients with T2DM showed that patients with OSA had lower insulin secretion reserve than those without OSA [79], but there was no adjustment for adiposity differences between groups.

**Mechanisms underpinning the relationship between OSA and dysglycemia**

Despite the strong association between OSA and dysglycemia, the underlying mechanisms are not clear. There are, however, several possible mechanisms that link OSA to dysglycemia, IR, and β-cell dysfunction, such as changes in sleep architecture, sympathetic overactivation, increased inflammatory cytokines, hormonal changes, increased oxidative stress, and the development of nonalcoholic fatty liver disease (NAFLD) (Figure 22.2).

Intermittent hypoxia is an important component of OSA and may contribute to much of its pathologic consequences. Intermittent hypoxia for as little as 5 hours in healthy volunteers can reduce insulin sensitivity without compensatory increase in insulin secretion, suggesting an impact on β-cell function as well [80]. Intermittent hypoxia is associated with increased hypoxia-inducible factor-1 (HIF-1) which occurs either secondary to hypoxia itself [81] or to oxidative stress [82]. HIF-1 upregulates sterol regulatory element-binding protein (SREBP)-1 [83], which is associated with increased lipid biosynthesis and insulin resistance [84]. HIF is also involved in systemic inflammation [85]. Interestingly, in mice with partial deficiency of HIF-1, intermittent hypoxia does not result in IR [3,83].

In addition to intermittent hypoxia, OSA is associated with changes in sleep architecture such as reduction in slow wave sleep and sleep quality resulting in excessive daytime sleepiness. These changes in sleep architecture and quality have
been associated with a reduction in insulin sensitivity and dysglycemia [86].

OSA is associated with many hormonal changes that can affect glucose metabolism. These include activation of the hypothalamic pituitary adrenal (HPA) axis and suppression of the GH axis and IGF-1; some of which can be reversed with CPAP treatment [87,88]. Gherlin has been shown to be higher in patients with OSA, which is reduced by CPAP treatment [89]. Catecholamines are also elevated in patients with OSA and are lowered by CPAP [90].

OSA is also associated with changes in adipokines. Adiponectin levels correlate negatively with the OSA severity independent of age, BMI, and visceral fat volume [91]. Several other studies showed lower adiponectin levels in OSA patients, although short CPAP treatment did not seem to reverse this trend [92,93]. Inversely to adiponectin, leptin levels were shown to be higher in obese subjects with OSA compared to age and BMI-matched obese subjects without OSA (p < 0.05) [50]. There are several studies that similarly showed increased leptin levels in patients with OSA [49].

Sympathetic overactivation plays an important role in the regulation of glucose and fat metabolism and the development of T2DM [94]. OSA is associated with increased sympathetic activity [92]. It is likely that both the recurrent hypoxia [95] and recurrent arousals [96] contribute to the activation of the sympathetic system. OSA is also associated with elevated inflammatory cytokines such as IL-6, TNF-α, and NF-κB [92].

OSA may also be a risk factor for the developing of histologically proven nonalcoholic fatty liver disease (NAFLD) and for progressing to NASH [66,97]. Nocturnal desaturations were found to be associated with hepatic inflammation, hepatocyte ballooning, and liver fibrosis [66]. Another study also found that subjects with histologic NASH had significantly lower desaturation, lower mean nocturnal oxygen saturation, and higher AHI compared with non-NASH controls [97]. A randomized controlled trial showed that four weeks of CPAP had no impact on liver enzymes in patients with moderate to severe OSA [98]; it must be noted, however, that liver enzymes are not a sensitive measure of NAFLD/NASH.

Recurrent hypoxia and mitochondrial dysfunction in OSA result in the formation of reactive oxygen species (ROS) which leads to cellular and DNA damage and oxidative stress [99]. Many studies support OSA as a cause of oxidative stress [92]. Repetitive episodes of re-oxygenation following hypoxia, as seen in OSA, simulate ischemia–reperfusion injury which results in the generation of ROS [100]. ROS levels have been shown to be 2–3 times higher in patients with OSA compared to healthy controls [99]. In addition, multiple studies have shown increased oxidized lipids, DNA, and carbohydrates in patients with OSA and animals exposed to intermittent hypoxia [99].

All the afore-mentioned mechanisms can impact on IR and/or β-cell function resulting in impaired glucose metabolism and eventually T2DM.

OSA and T2DM

OSA and incident T2DM

Several studies have examined whether having OSA increases the individual's risk of developing T2DM, in particular whether obesity is a major risk factor for both conditions and whether OSA has been associated with IR and prediabetes as described earlier. Several cross-sectional studies found a higher prevalence of T2DM in patients with OSA despite adjusting for confounders, particularly age and obesity; these studies are reviewed in [101]. Whether OSA is a predictor of T2DM has been examined in a small number of longitudinal studies that used a variety of methods to diagnose OSA (from symptoms to polysomnography), and to diagnose T2DM (from self-reported to OGTT); these studies are summarized in Table 22.2. A recent meta-analysis of studies that used objective measures to diagnose OSA found that moderate to severe OSA was associated with increased risk of developing T2DM (RR 1.63; 95% CI 1.09–2.45) [102].

OSA prevalence in T2DM

Given the high prevalence of glycemic abnormalities in patients with OSA, it is not surprising that OSA is very common in patients with T2DM. However, there is significant variation in OSA prevalence between studies due to differences in population characteristics (primary vs. secondary care, long vs. short diabetes duration, ethnicity, obesity, and so on), and differences in the methods and criteria used to diagnose OSA. These studies are summarized in Table 22.3.

As a result of the high prevalence of OSA in patients with T2DM, the International Diabetes Federation (IDF) recommended screening for OSA in this high-risk population [108], although appropriate validated screening methods in patients with T2DM are still lacking.

OSA and glycemic control in T2DM

Since OSA is associated with IR and possibly β-cell dysfunction, and increased inflammation and oxidative stress (see later for more details), it is logical to hypothesize that OSA leads to worsening glycemic control in patients with T2DM. The main difficulty in OSA-related research is to differentiate the impact of OSA from that of obesity. A small number of studies assessed this issue, and they demonstrated that OSA and OSA severity are associated with poorer glycemic control (both HbA1c and fasting plasma glucose) and glycemic variability after multivariable adjustments for several confounders, such as age, sex, race, BMI, number of diabetes medications, level of exercise, years of diabetes and total sleep time in some studies [117,120–123]. These studies were relatively small (n = 31–92). The adjusted mean increase in HbA1c between patients with and without OSA varied from 0.7% to 3.69% depending on the OSA severity. One study, however, did not show an association between OSA and glycemic control [112]; but in this study only 22% of participants had full polysomnography and the duration of the sleep study was just 4 hours [77]. A recent study showed...
### Table 22.2: Prospective studies that examined the association between OSA (or surrogate markers like snoring) and incident T2DM. Patients in these studies were free from T2DM at baseline

<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Population</th>
<th>Follow up (year)</th>
<th>OSA diagnosis</th>
<th>T2DM diagnosis</th>
<th>Results</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elmasry et al. [32] 2000</td>
<td>Sweden</td>
<td>2668 (all men, aged 30–69 years)</td>
<td>10</td>
<td>Questionnaire</td>
<td>Questionnaire</td>
<td>DM development: 5.4% vs. 2.4% with vs. without habitual snoring respectively ($p &lt; 0.001$). 13.5% vs. 8.6% for obese snorers vs. obese nonsnorers, respectively ($p = 0.17$). OR (95% CI): 7.0 (2.9–16.9) vs. 5.1 (2.7–9.5) for obese snorers vs. obese nonsnorers</td>
<td>Adjusted for: age, weight gain, smoking, alcohol dependence, and physical inactivity</td>
</tr>
<tr>
<td>Al-Delaimy et al. [103] 2002</td>
<td>USA</td>
<td>69852 (all female nurses, all nondiabetics at baseline)</td>
<td>10</td>
<td>Questionnaire</td>
<td>Self-reported and questionnaire</td>
<td>Snoring was associated with risk of diabetes (occasional vs. nonsnoring, RR 1.48; 95% CI 1.29–1.70; regular vs. nonsnoring, RR 2.25; 95% CI 1.91–2.66)</td>
<td>Adjusted for age and BMI</td>
</tr>
<tr>
<td>Reichmuth et al. [104] 2005</td>
<td>USA</td>
<td>1387</td>
<td>4</td>
<td>PSG</td>
<td>FPG physician diagnosis</td>
<td>Adjusted OR for developing DM with an AHI of 15 or more: 1.62 (95% CI 1.07–2.45)</td>
<td>Adjusted for age, sex, and body habitus</td>
</tr>
<tr>
<td>Marshall et al. [105] 2009</td>
<td>Australia</td>
<td>295 (population-based)</td>
<td>4</td>
<td>PSG</td>
<td>Physician diagnosis or FPG</td>
<td>Moderate to severe OSA was an independent risk factor for incident diabetes (OR = 13.45; 95% CI 1.59–114.11)</td>
<td>Adjusted for age, gender, waist circumference, BMI, BP, and HDL. Large 95% CI due to small number of incident diabetes</td>
</tr>
<tr>
<td>Botros et al. [106] 2009</td>
<td>USA</td>
<td>544</td>
<td>2.7</td>
<td>PSG</td>
<td>Physician diagnosis and FPG</td>
<td>Adjusted HR per quartile of OSA severity = 1.43 (95% CI 1.10–1.86)</td>
<td>OSA defined as AHI ≥ 8. Adjusted for age, sex, race, BMI, baseline fasting glucose, weight change</td>
</tr>
<tr>
<td>Celen et al. [107] 2010</td>
<td>Sweden</td>
<td>168 (without DM from a sleep clinic)</td>
<td>16</td>
<td>PSG</td>
<td>Physician diagnosis</td>
<td>Incident T2DM: 15.5% vs. 24.6% ($p = 0.02$). OSA was a predictor of incident T2DM in women (OR 11.8; 95% CI 1.1–121.7, $p = 0.04$) but not men after adjustment</td>
<td>Adjusted for age and BMI</td>
</tr>
<tr>
<td>Lindberg et al. [42] 2012</td>
<td>Sweden</td>
<td>141 (all men without DM)</td>
<td>11</td>
<td>Overnight respiratory monitor</td>
<td>OGTT</td>
<td>ODI &gt; 5 was a predictor of developing DM (adjusted OR 4.4; 95% CI 1.1–18.1)</td>
<td>Adjusted for age, BMI, change in BMI, HTN, and CPAP treatment</td>
</tr>
</tbody>
</table>
### Table 22.3 Summary of studies that examined OSA prevalence in patients with T2DM

<table>
<thead>
<tr>
<th>Study</th>
<th>Population</th>
<th>Samples size</th>
<th>OSA diagnosis</th>
<th>OSA prevalence</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brooks et al. [109]</td>
<td>Australia; BMI &gt; 35</td>
<td>31</td>
<td>Ambulatory sleep monitoring</td>
<td>70% moderate to severe OSA</td>
<td>Sample was selected from a larger population based on OSA symptoms</td>
</tr>
<tr>
<td>Elmasry et al. [51]</td>
<td>Hypertensive men, 21% had diabetes. Age 61.4 (8.0), BMI 29.3 (4.5)</td>
<td>116</td>
<td>PSG; OSA defined as AHI ≥ 20</td>
<td>36% in the diabetes group</td>
<td>The sample that had PSG was chosen based on questionnaires</td>
</tr>
<tr>
<td>Resnick et al. [110]</td>
<td>A subgroup from the Sleep Health Heart study</td>
<td>470</td>
<td>PSG; OSA defined as RDI ≥ 5. Moderate to severe RDI ≥ 15</td>
<td>OSA prevalence 57.8%. Moderate to severe 23.8%</td>
<td>Self-reported diabetes diagnosis or use of medications</td>
</tr>
<tr>
<td>West et al. [111]</td>
<td>All male, mixed primary and secondary care populations, UK. Age 61.2 (9.7), BMI 29.6 (5.4)</td>
<td>1676</td>
<td>Oximetry; OSA defined as ODI &gt; 10</td>
<td>23%</td>
<td>Population screened by questionnaires. A subgroup was selected for oximetry</td>
</tr>
<tr>
<td>Einhorn et al. [112]</td>
<td>Consecutive adults with T2DM from a diabetes clinic in the USA</td>
<td>330</td>
<td>Single-channel device that measured nasal airflow. OSA defined as AHI ≥ 10</td>
<td>48% In moderate to severe OSA: 36%</td>
<td></td>
</tr>
<tr>
<td>Laaban et al. [113]</td>
<td>Consecutive hospitalized patients with poorly controlled T2DM</td>
<td>303</td>
<td>Overnight ventilatory polygraphic study. OSA defined as RDI ≥ 5, Moderate to severe RDI ≥ 15</td>
<td>OSA: 63% Moderate to severe: 29%</td>
<td></td>
</tr>
<tr>
<td>Foster et al. [114]</td>
<td>Community-based population from USA (19.1% Afro-Caribbean). Age 61.3 (6.5), BMI 36.5 (5.8)</td>
<td>306</td>
<td>PSG; OSA defined as AHI ≥ 5. Moderate to severe OSA: AHI ≥ 15/30</td>
<td>86% 30.5% for moderate OSA: 22.6% for severe OSA</td>
<td>Only overweight or obese individuals were included</td>
</tr>
<tr>
<td>Lam et al. [115]</td>
<td>Randomly selected patients from a teaching hospital diabetes clinic in China. Age 57.3 (9.3), BMI 26.0 (4.6)</td>
<td>165</td>
<td>PSG; OSA defined as AHI ≥ 5. Moderate to severe OSA: AHI ≥ 15</td>
<td>53.9% had OSA; 32.7% had moderate to severe OSA</td>
<td>Patients with RRT were excluded</td>
</tr>
<tr>
<td>Schober et al. [116]</td>
<td>Secondary care sample from Germany</td>
<td>498</td>
<td>Multichannel respiratory device. OSA defined as AHI ≥ 15</td>
<td>37.4%</td>
<td>This study also included patients with type 1 diabetes; OSA prevalence 10.3%</td>
</tr>
<tr>
<td>Pillai et al. [117]</td>
<td>Consecutive patients from secondary care diabetes obesity clinic in the UK</td>
<td>52</td>
<td>PSG</td>
<td>58%</td>
<td>Participants had risk factors for OSA</td>
</tr>
<tr>
<td>Tahrani et al. [118]</td>
<td>Randomly selected patients from secondary care in UK, 45% are South Asians. Age 57 (12), BMI 34.4 (30.9–39.5)</td>
<td>234</td>
<td>Multichannel cardiorespiratory device. OSA defined as AHI ≥ 5</td>
<td>65% Moderate to severe OSA: 26%</td>
<td>Patients with RRT were excluded. Prevalence was not the primary outcome of the study</td>
</tr>
<tr>
<td>Heffner et al. [119]</td>
<td>Case notes study from primary care in the US. Age 64 (14.1), BMI 33.7 (8.3)</td>
<td>16066</td>
<td>Physician diagnosis</td>
<td>18% of known OSA; 23% had OSA amongst obese patients</td>
<td>This study did not screen for OSA, it simply reports the prevalence of known OSA, hence the lower prevalence than other studies</td>
</tr>
</tbody>
</table>

PSG, polysomnography; RRT, Renal Replacement Therapy.
that following adjustment for age, gender, obesity, smoking status, hypertension and antihypertensives, AHI correlated with HbA1c in patients with prediabetes but not in those with T2DM or normoglycemia, whilst the lowest oxygen saturation correlated with HbA1c across all groups [124].

Despite the constant finding of an association between OSA and glycemic control in patients with T2DM, the impact of CPAP on glycemic control has not been evaluated widely and the results were not consistent (Table 22.4). Several studies [71,109,125–129] have evaluated CPAP treatment; of these, only one is a randomized clinical trial [71], with the rest being uncontrolled pre/post assessments (Table 22.4). The one controlled study showed no change in HbA1c after CPAP therapy for 3 months. The lack of positive effect could be attributed to the small study sample, the limited duration of follow-up, the suboptimal adherence to CPAP (3.6 hours per night) or a true lack of effect. In marked contrast, uncontrolled studies have shown improvements in insulin sensitivity [109,125], postprandial hyperglycemia [126], glycemic variability [129], or HbA1c [126,127] (Table 22.4). A recent meta-analysis found that CPAP treatment did not result in significant reductions in HbA1c (0.08 %; 95 % CI -0.26–0.42) in patients with T2DM [130]. However, new data showing that AHI only correlated with REM AHI and not non-REM AHI suggest that longer usage of CPAP beyond 4 hours is needed as most REM sleep occurs in the latter half of the night [131]. Indeed recent preliminary results showed that 1 week of in-laboratory (8 hours of sleep) CPAP treatment resulted in a decrease of 11.2 and 19.8 mg dL$^{-1}$ in the average 24-hour and post-breakfast glucose levels, respectively, with a 45% decrease in the dawn phenomenon as well [132]. However, enforcing 7–8 hours per night of CPAP usage in real life is a huge challenge. In addition, it is important to note that the diabetes duration of patients in this study was relatively short (3.2 years) unlike in previous studies; this is important as IR and β-cell dysfunction may be more reversible in such a population.

**OSA and diabetes-related complications**

Recently, there has been increasing interest in the association between OSA and diabetes-related complications. There is one cross-sectional published study regarding the association between OSA and macrovascular complications in T2DM, but several studies have examined the association between microvascular complications and OSA in T2DM. Most of these studies are cross-sectional in nature but more recently prospective studies proving causality have been published. Many of these studies fail to adjust for many of the possible confounders.

The hypothesis that OSA can cause micro- (as well as macro-) vascular complications in T2DM is based on the observation that OSA and intermittent hypoxia result in similar molecular consequences in patients without T2DM to those caused by hyperglycemia and result in diabetes-related complications; including increased oxidative stress and advanced glycation end-products, production and activation of the protein kinase C pathway and its molecular perturbations resulting in micro- and macrovascular disease and endothelial dysfunction (Figure 22.3) [118]. Recent work has shown that OSA in patients with T2DM is associated with increased oxidative stress and nitrosative stress and impaired microvascular and endothelial function despite adjustment for a wide range of confounders [118]. OSA has also been shown to be associated with poly (ADP ribose) polymerase (PARP) activation in patients with T2DM (Tahran, unpublished data).

One cross-sectional analysis from the Look AHEAD study showed that AHI is associated with stroke (adjusted OR 2.57; 95% CI 1.03–6.42), but there was no association with coronary artery disease [133]. Cardiovascular disease (CVD) in this study was self-reported.

Three studies have examined the association between OSA and diabetic retinopathy (DR). In the first study, Shiba et al. used a highly selected population of 219 Japanese patients who were undergoing vitreous surgery and used pulse oximetry to diagnose OSA [134]. Patients with proliferative DR had higher ODI compared to those with nonproliferative DR. After adjustment for age, HbA1c, and hypertension, higher oxygen saturations were found to be protective against proliferative DR [134]. The same investigation has also shown that OSA is associated with angle neovascularization in proliferative DR [135]. Another study included 118 men from primary and secondary care in the UK [136] and found that OSA was independently associated with DR and maculopathy after adjusting for age, BMI, diabetes duration, and hypertension [136]. Another cross-sectional study from the UK that included 226 patients recruited from a secondary care diabetes clinic and used multichannel respiratory devices to diagnose OSA, found that patients with OSA are 3–4 times more likely to have sight-threatening retinopathy, preproliferative/proliferative retinopathy or maculopathy after adjustment for a wide range of confounders, such as age, obesity, diabetes duration, glycemic control and medications amongst others [137]. Furthermore, a longitudinal follow-up of the last mentioned cross-sectional study showed that patients with OSA are more likely to develop preproliferative or proliferative DR despite adjustment for confounders (OR 6.6; 95% CI 1.2–35.1, $p = 0.03$) [137]. In an uncontrolled, hypothesis-generating study of 28 patients with clinically significant macular edema, 6 months of treatment with CPAP was associated with improvement in visual acuity without an impact on macular edema [138].

The association between OSA and diabetic peripheral neuropathy (DPN) has been examined in one cross-sectional study that included South Asians and white Caucasians with T2DM recruited from the diabetes clinic of a secondary care hospital in the UK. This study used portable multichannel respiratory devices to diagnose OSA, the Michigan Neuropathy Screening Instrument to diagnose DPN, and the 10 g monofilament perception to diagnose foot insensitivity. Patients with OSA were three times (OR 2.82; 95% CI 1.44–5.52) and four times (OR 3.97; 95% CI 1.80–8.74) more likely to have DPN and...
Table 22.4 Summary of studies that examined the impact of CPAP treatment on glycemic control in patients with T2DM

<table>
<thead>
<tr>
<th>Study</th>
<th>Population</th>
<th>Study design</th>
<th>Control group</th>
<th>Matching or confounder adjustment</th>
<th>Outcome measure</th>
<th>Duration</th>
<th>CPAP usage</th>
<th>Effectiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brooks et al. [109] 1994</td>
<td>n = 10 Australia Obese T2DM (BMI &gt; 35)</td>
<td>Pre and post</td>
<td>No</td>
<td>No</td>
<td>Glucose disposal during hyperinsulinemic euglycemic clamp</td>
<td>4 months</td>
<td>NR</td>
<td>+</td>
</tr>
<tr>
<td>Harsch et al. [125] 2004</td>
<td>n = 9 Severe OSA BMI 37.3 (5.6) HbA1c 6.4 (0.7)%</td>
<td>Pre and post</td>
<td>No</td>
<td>No</td>
<td>ISI established by euglycemic hyperinsulinemic clamp</td>
<td>3 months</td>
<td>5.8 h/n</td>
<td>-</td>
</tr>
<tr>
<td>Babu et al. [126] 2005</td>
<td>n = 25 Severe OSA BMI 42.7 (8.7) HbA1c 8.3 (2.2) Diabetes duration 8.6 (6.3)</td>
<td>Pre and post</td>
<td>No</td>
<td>No</td>
<td>72 h CGMS and HbA1c</td>
<td>3 months</td>
<td>4.2 h/n</td>
<td>+</td>
</tr>
<tr>
<td>Hassaballa et al. [127] 2005</td>
<td>n = 38 Severe OSA BMI 42.9 (9.5) HbA1c 7.8 (1.4)% Diabetes duration 8.6 (6.3)</td>
<td>Pre and post</td>
<td>No</td>
<td>No</td>
<td>HbA1c</td>
<td>4 months</td>
<td>4 h/n</td>
<td>+</td>
</tr>
<tr>
<td>West et al. [71] 2007</td>
<td>n = 42 Age 57.8 (10.4) BMI 36.6 (4.9) HbA1c 8.5 (1.8)</td>
<td>RCT</td>
<td>Yes</td>
<td>N/A</td>
<td>HbA1c</td>
<td>3 months</td>
<td>3.6 h/n</td>
<td>-</td>
</tr>
<tr>
<td>Dawson et al. [128] 2008</td>
<td>n = 20 Moderate to severe OSA Age 59.8 (10.2) BMI 39.6 (8.0) Diabetes duration 9.8 (7.7) HbA1c 7.2%</td>
<td>Pre and post</td>
<td>No</td>
<td>No</td>
<td>CGMS HbA1c</td>
<td>41 days</td>
<td>5.8 h/n</td>
<td>+ for glucose variability - for HbA1c</td>
</tr>
<tr>
<td>Pallayova et al. [129] 2008</td>
<td>n = 14 Severe OSA Age 54 (6) BMI 37.4 (6.3) Diabetes duration 3.7 (1.5) HbA1c 7.48 (0.92)%</td>
<td>Pre and post</td>
<td>No</td>
<td>No</td>
<td>CGMS</td>
<td>NR</td>
<td>NR</td>
<td>+</td>
</tr>
</tbody>
</table>

CGMS, continuous glucose monitoring system; h/n, hours per night; n, number; NR, not reported.
Figure 22.3 The postulated mechanisms linking OSA to DPN (and microvascular complications). AGE, advance glycation end-products; HTN, hypertension; PKC, protein kinase C; ROS, reactive oxygen species; RNS, reactive nitrogen species. For more details please refer to the text. Source: Adapted from Tahrani et al. [118].

foot insensitivity, respectively, compared to those without OSA [118]. OSA patients were also more likely to have reduced vibration perception and absent ankle jerk reflex [118]. OSA and nocturnal hypoxemia severity were associated with DPN and its severity [118]. Unpublished data from the same study showed that OSA was also associated with lower intra-epidermal nerve fiber density in skin biopsies (Tahrani, unpublished data) suggesting that OSA may cause both large as well as small nerve fiber dysfunction.

The association between OSA and diabetic nephropathy has been examined in two studies. In one study of 237 patients, snoring (as a surrogate marker of OSA) was found to be independently associated with microalbuminuria [139]. Another study showed that OSA was associated with diabetic nephropathy (defined as microalbuminuria, microalbuminuria and/or reduced eGFR) in patients with T2DM after adjustment for a wide range of confounders; patients with end-stage renal disease were excluded from this study [140]. A 2.5 year follow-up of the latest study showed that the decline in eGFR was greater in patients with T2DM and OSA compared to those with OSA alone (−6.8% [−16.1–2.2%] vs. −1.6% [−7.7–5.3%], p = 0.002) and that OSA was a predictor of study-end eGFR despite adjustment for confounders [140].

OSA is also associated with autonomic neuropathy and increased sympathetic activity [141]. It is likely that both recurrent hypoxia [95] and recurrent arousals [96] contribute to the activation of the sympathetic system. Obese subjects with autonomic neuropathy develop more frequent and more prolonged hypopnea/apnea in comparison to obese subjects without autonomic dysfunction whether or not they had T2DM [142].

Other OSA comorbidities

Road traffic accidents
Several cross-sectional studies using driving simulators showed worse driving performance and increased risk of road traffic accidents in patients with OSA [143]. In a sample of 913 employed adults whose motor vehicle accident history was obtained from a state-wide database covering the period 1988–1993, men with OSA (AHI ≥ 5) were more likely to have at least one accident in 5 years compared to those without OSA (age and miles driven adjusted OR 4.2 for AHI 5–15, and 3.4 for AHI > 15) [10,144]. Men and women combined with AHI > 15 (vs. no OSA) were significantly more likely to have multiple accidents in 5 years (OR 7.3) [10,144]. Similar results were found in another case-control study from Spain [145]. Interestingly, neither of these two studies showed a relationship between reported sleepiness and the road traffic accidents nor was there a dose relation between OSA severity and the likelihood of involvement in an accident.

Cardiovascular mortality and morbidity
OSA has been associated with sustained hypertension and lack of the normal nocturnal dipping of blood pressure (BP). Nondipping of BP in OSA patients was examined prospectively in a subsample of 328 adults enrolled in the Wisconsin Sleep Cohort Study who completed 2 or more 24-hour ambulatory BP studies over an average of 7.2 years [146]. After adjustment for many confounders, the adjusted OR (95% CI) of incident systolic nondipping for baseline AHI 5–14.9 and ≥15, vs. AHI < 5, were 3.1 (1.3–7.7) and 4.4 (1.2–16.3), respectively [146]. In a cross-sectional analysis of 6132 middle-aged and older persons
The incidence of CVD compared to those untreated (56.8% vs. 6.7%, confounders [149]. CPAP treatment was associated with lower of CVD (OR 4.9; 95% CI 1.8–13.6) after adjustment for variables.

Several cross-sectional and case-control studies showed an association between OSA and cardiovascular disease (CVD) [10]. Three prospective studies used more accurate methods to diagnose OSA (i.e., polysomnography) [149–151]. In a study of a 182 consecutive middle-aged men free of CVD at baseline who were followed for 7 years, the incidence of CVD was 36.7% of patients with OSA vs. 6.6% of subjects without OSA (p < 0.001) [149]. OSA was an independent predictor of CVD (OR 4.9; 95% CI 1.8–13.6) after adjustment for confounders [149]. CPAP treatment was associated with lower incidence of CVD compared to those untreated (56.8% vs. 6.7%, p < 0.001) [149]. In another prospective study in which men with OSA were followed for a mean of 10.1 years, patients with untreated severe OSA had a higher incidence of fatal CVD and nonfatal CVD than did untreated patients with mild-moderate OSA, simple snorers, patients treated with CPAP, and healthy participants [150]. After adjustment for confounders, untreated severe OSA significantly increased the risk of fatal OSA (OR 2.87; 95% CI 1.7–4.9, 5.0–14.9, and ≥15.0, respectively (p = 0.002 for the trend) [148].

OSA should be treated promptly and the aim of treatment is to reduce the morbidity and mortality associated with this condition. Weight loss and positional treatment (i.e., avoiding the position in which most episodes occur, which is usually the supine position) are important aspects of treatment. As with all obesity-related disorders, weight loss (regardless of the means) can result in significant improvements in OSA. In a randomized controlled trial of intensive lifestyle intervention in 264 patients with OSA and T2DM (the Sleep AHEAD study), weight loss of 11 kg on average in the treatment group resulted in a reduction in the AHI of about 10 events per hour [157]. Similar results were found in a study of men with OSA, in which 10% weight loss resulted in improvements in the RDI by about 16 events per hour [158]. Weight loss after bariatric surgery has also been associated with significant improvements in OSA severity [159].

Mandibular advancement devices (MAD) are effective in treating patients with mild to moderate OSA. They work by pulling the tongue forward or by moving the mandible and soft palate anteriorly, enlarging the posterior airspace, which results in opening and increasing the airway size. MAD are considered as a second-line treatment for patients with mild to moderate OSA who cannot tolerate CPAP [160]. Surgery has a limited role in patients with OSA and produces variable results [160]. If the patient has upper airway obstruction (such as tonsils or tumors) then surgery is the most important aspect of the treatment, otherwise its role is limited and it is usually associated with significant side effects [160].

CPAP is the mainstay of treatment for patients with OSA. CPAP works by providing a “pneumatic splint” by delivering an intraluminal pressure that is positive with reference to the atmospheric pressure and by increasing lung volumes [161]. CPAP treatment has been shown to reduce AHI, reduce BP, improve sleepiness, improve quality of life, improve cognitive function, and reduce motor vehicle accidents [161]. Furthermore, evidence from an observational study suggests that CPAP treatment reduces the risk of cardiovascular events [150]. An in-depth review on CPAP, its technical aspects, evidence behind its use, and its complications can be found elsewhere in [161].
Impact of T2DM on OSA

Although the bulk of this chapter describes the impact of OSA on glucose metabolism and T2DM, this association could be directional and having IR or T2DM may have an impact on pre-existing OSA and result in the development of “new” OSA or CSA.

In a study of 3565 participants, who were followed up for a 6-year period, the presence of a history of witnessed apneas was an independent predictor of incident sleep apnea [162]. After adjustment for age, sex, and waist circumference, the standardized OR for incident sleep apnea for HOMA-IR was 1.31 (1.13–1.51). This suggests that not only can sleep apnea result in dysglycemia as shown by longitudinal studies, but that pre-existing dysglycemia also predicts the development of sleep apnea. This may be related in part to the loss of upper airway innervations or the autonomic dysfunction that can occur in patients with dysglycemia; autonomic neuropathy has been implicated in the central respiratory centre response to hypercapnic stimulus [163] and may result in changes in respiratory control resulting in CSA as seen in other conditions such as multisystem atrophy (Shy–Drager syndrome) [164]

CSA has indeed been shown to be common in patients with T2DM. In a subgroup analysis of the Sleep Heart Health Study, there were significant differences in RDI, sleep stages, central apnea index, and periodic breathing between patients with and without DM. However, most of these differences lost their statistical significance after adjusting for confounders with the exception of percent time in REM sleep and prevalence of periodic breathing [110]. Similarly, Sanders and colleagues found a greater prevalence of CSA in patients with DM compared to those without (3.8% vs. 1.8%, DM vs. non-DM patients, respectively, \( p = 0.002 \)) [165].

The natural history of OSA in patients with T2DM and the impact of T2DM on pre-existing OSA is currently unknown, but the ongoing Sleep AHEAD study will be able to answer this question.

Summary and future directions

OSA is a very common medical disorder that is associated with significant morbidity and mortality. OSA is associated with obesity, hypertension, insulin resistance, and possibly impaired \( \beta \)-cell function. OSA is associated with several mechanisms that lead to insulin resistance and \( \beta \)-cell dysfunction including intermittent hypoxia, inflammation, sympathetic activation, activation of the hypothalamic pituitary adrenal axis, reduction in adiponectin, increased leptin, and the development of nonalcoholic steatohepatitis. As a result, dysglycemia is very common in patients with OSA.

OSA is associated with worse glycemic control in patients with T2DM but CPAP treatment does not seem to improve glycemic control. The impact of CPAP treatment on other metabolic parameters in patients with T2DM including blood pressure and dyslipidemia are also needed. OSA is associated with increased oxidative and nitrosative stress and impaired microvascular and endothelial function in patients with diabetes, hence it is not surprising that OSA is associated with diabetes-related microvascular complications. However, prospective and interventional studies are required to confirm causality. In addition, T2DM is associated with CSA, and not only OSA, and this seems to be related to autonomic dysfunction.

The field of OSA in patients with T2DM is still in its infancy and most of the work in regard to OSA in patients with T2DM has focused on glycemic control, it is time to address the impact of OSA on other aspects of T2DM.

Declaration

Dr. Abd Tahran is a clinician scientist supported by the UK National Institute for Health Research. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research, or the Department of Health.

References


119 Drager LF, Queiroz EL, Lopes HF, et al.: Obstructive sleep apnea is highly prevalent and correlates with impaired glycemic control in consecutive patients with the metabolic syndrome. Journal of the Cardiometabolic Syndrome 2009;4:89–95.


SECTION V

PATHOGENESIS
Key points

- Aside from insulin’s role to regulate glucose metabolism via effects on skeletal muscle, liver, and adipose tissue, it has effects on numerous other tissues (e.g. endothelium and immune system) and risk pathways and, consequently, insulin resistance is linked to many risk factor perturbations.
- The simplest clinically recognizable and most dominant feature of insulin resistance is presence of obesity, whether measured by body mass index (BMI) or elevated waist circumference.
- The characteristic insulin resistance lipid pattern is one of elevated triglyceride and lower HDL-cholesterol concentrations and these features, together with obesity and elevated blood pressure, have formed the cornerstone of most versions of metabolic syndrome diagnostic criteria.
- More recently, elevated liver fat, recognizable in some by elevated ALT and GGT levels, has emerged as a key and clinically important feature of insulin resistance in many individuals.
- Insulin resistance is also linked variably to alterations in adipokines, vascular dysfunction and immune regulation, and although the clinical relevance of these associations remain to be fully elucidated, novel anti-inflammatory agents are being trialed to lower glucose levels in diabetes.
- An attempt to bring the insulin resistance syndrome to wide clinical utility was made by introducing simpler metabolic syndrome criteria into the ATP III guidelines.
- Whilst the afore-mentioned metabolic syndrome was demonstrated to be clearly associated with cardiovascular events, prospective research eventually showed it did not improve cardiovascular risk prediction.
- Nevertheless the insulin resistance or metabolic syndrome is, arguably, conceptually useful and it may have led more clinicians to consider (and target) glycemia and adiposity parameters in their patients.
- Future research in this area should be better directed at determining causes and consequences of insulin resistance by harnessing the power of genetics; so-called Mendelian randomization studies.
- Finally, the strongest risk factor for insulin resistance remains excess adiposity and, as such, preventing or targeting excess weight gain must remain the major goal in public health to reduce the incidence of its clinical sequelae, in particular, type 2 diabetes.

What is insulin resistance?

Broadly, insulin resistance can be defined as an abnormal biologic response to insulin; insulin, whether endogenous or exogenous in origin, has limited ability to reverse a hyperglycemic metabolic state. Thus a person with insulin resistance is, almost inevitably, progressing towards developing frank type 2 diabetes (T2DM) if an intervention (usually lifestyle) is not implemented. Due to the very close link between diabetes and insulin resistance, no formal clinical definition of insulin resistance has emerged. We will discuss the potential clinical use of the insulin resistance syndrome and the metabolic syndrome later in the chapter. For now, we focus on investigation of insulin resistance as a useful entity for research concepts, and in particular discuss its biologic consequences by describing its associated risk factor perturbations.

Multiple roles of insulin

To understand the link between many novel parameters and insulin resistance, and thereby their association with T2DM, it is important to appreciate that insulin imparts its effects not only on skeletal muscle but also on many tissues including adipose, liver, endothelium, and immune cells. Thus, insulin is relevant not simply for glucose uptake and metabolism, but also for (i) suppression of free fatty acid (FFA) release from adipose tissue, (ii) limiting hepatic triglyceride synthesis, (iii) helping to maintain endothelial homeostasis, (iv) regulating thrombotic cascades, and (v) potentially playing a role in regulating inflammatory cascades (Figure 23.1). With increasing adiposity, fat cells become enlarged and less responsive to insulin, that is, they become insulin resistant. The subsequent excessive release of FFAs into the portal circulation in part drives excess hepatic fat accumulation and triglyceride synthesis (leading to elevated circulating triglyceride levels and associated changes in circulating lipids) and muscle fat accumulation. Excess fat in the latter two organs is considered to enhance their insulin resistance;
Chapter 23

Adipose tissue
Muscle
Liver
Pancreas
Insulin
Glucose uptake
↑ FFAs
↑ TG > ↓ HDL-C
ALT, GGT, CRP, PAI-1

Cytokines?

monocytes

Figure 23.1 Multiple roles of insulin. The figure illustrates the relevance of insulin action not simply to skeletal muscle glucose uptake, but also adipose, liver, endothelium, and immune cell function. As fat cells enlarge with greater obesity and become more insulin resistant, they enhance (e.g. IL-6) or suppress (e.g. adiponectin) release of many agents into the circulation. Similarly, a subclinical “fatty liver,” dysfunctional endothelium, and irritated immune cells in insulin resistance release an array of molecules into the circulation. As a result, the circulating concentrations of such parameters give indications as to the degree of insulin resistance and also therefore associate with future risk for type 2 diabetes.

specifically, excess fat in the liver limits glucose storage (in the form of glycogen) but enhances glucose synthesis via gluconeogenesis, whereas excess fat in muscle lessens glucose uptake. The combined effects thus promote hyperglycemia. This hyperglycemia can be offset for many years by an increase in pancreatic insulin secretion, and hence hyperinsulinemia accompanies insulin resistance. In persons who develop T2DM, the pancreas has eventually become “exhausted”—it can no longer produce sufficient insulin to counteract the hyperglycemic drive—and glucose concentrations rise into the diabetic range.

How do we measure insulin resistance?

The gold standard method for measuring insulin resistance focuses on maintaining a normal blood glucose level in a hyperinsulinemic state—the so-called euglycemic hyperinsulinemic clamp. This technique involves infusing insulin at supraphysiologic rates based on body size, while also infusing glucose at the necessary rate to maintain euglycemia. The use of stable isotopes provides further information on the relevant roles of the liver and peripheral tissue (muscle and adipose tissue). When a steady state is achieved, the glucose infusion rate is essentially a measure of the ability of all tissues to respond to insulin and take up glucose for a standard insulin dose [1]. Thus, the global sensitivity to insulin is comparable between patients or participants. There are several different specific methodological variations of this approach to measure insulin resistance, each with strengths and weaknesses, although the aim is always similar. The major drawback of this gold-standard approach is its labor-intensive and time-consuming nature. Thus, it is not a practical test for routine screening or for large epidemiologic studies. Other dynamic tests to measure sensitivity to insulin are available but are infrequently used in clinical settings (Table 23.1).

Simple characteristics associated with insulin resistance

If one looks at any reputable risk score for the prediction of T2DM, certain parameters always appear—age, gender, adiposity measures, family history of T2DM, and ethnicity. Of all of these, rising obesity levels are strongly associated with insulin resistance whatever the other characteristics of an individual and, as such, fat accumulation is the primary risk factor that drives increasing insulin resistance at a population level. That noted, the other risk factors are also of importance in considering an individual’s risk. For example, an older Asian male with a family history of T2DM will have greater insulin resistance for a given weight than younger white females without a family history of diabetes. In other words, nonmodifiable characteristics (age, sex, ethnicity, family history) interact with adiposity to determine degree of insulin resistance and
### Table 23.1 Common methods of measurement of insulin resistance

<table>
<thead>
<tr>
<th>Method name</th>
<th>Method summary</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Direct model techniques</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Euglycemic hyperinsulinemic clamp (EUC)</td>
<td>Steady i.v. insulin infusion in one arm, glucose infusion in the other to achieve “clamped” euglycemic state.</td>
<td>The gold-standard method. Radio-labeled glucose tracer can be used to investigate tissue-specific glucose disposal. Hyperinsulinemic-hyperglycemic clamp modification estimates ( \beta )-cell function.</td>
<td>Labor intensive, time-consuming, and requires experienced operators. Test will run from early morning to early afternoon. Multiple tests to determine euglycemia achieved. Ideal insulin infusion rate varies according to study cohort. Hypoglycemia may occur.</td>
</tr>
<tr>
<td>Insulin sensitivity test (IST)</td>
<td>Infusion of defined glucose load and steady-state insulin to a single arm over (~3) h. Somatostatin can also be infused to limit the effect of endogenous insulin. Glucose concentration from contralateral arm every 10 min in last 30 min determines SI.</td>
<td>Highly reproducible. Easier to achieve steady state and shorter method than EUC. Fewer blood tests.</td>
<td>Still impractical for clinical setting/large studies. Hypoglycemia possible in insulin-sensitive individuals.</td>
</tr>
<tr>
<td>Insulin tolerance test (ITT)</td>
<td>Simplified version of IST. Measures fall in glucose for 15 min after i.v. bolus of insulin.</td>
<td>Short test. Little opportunity for interference from endogenous factors.</td>
<td>Only measures skeletal muscle IR. Increased risk of hypoglycemia.</td>
</tr>
<tr>
<td><strong>Indirect “minimal” models</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequently sampled IV glucose tolerance tests (FSIVGGT)</td>
<td>Several variations of this model exist. Glucose is infused (with or without insulin) and glucose is measured on multiple occasions over (~3) h. SI is calculated from a curve fit observing glucose disappearance.</td>
<td>Slightly easier than EUC. SI often correlates very well with EUC.</td>
<td>Many assumptions made about dynamics of glucose disposal (no steady state).</td>
</tr>
<tr>
<td>Continuous infusion of glucose with model assessment (CIGMA)</td>
<td>Like ITT, a bolus infusion of glucose is administered. Blood samples for glucose and insulin measures are drawn at a few time points after around 1 h. A mathematical model derives SI.</td>
<td>Not as time-consuming as FSIVGGT.</td>
<td>Not well validated.</td>
</tr>
<tr>
<td>Oral glucose tolerance test (OGTT)</td>
<td>This is a routine test for the clinical diagnosis of diabetes when simpler blood tests are ambiguous. Often involves a 75 g or 100 g dose of glucose, and measuring glucose and insulin at frequent intervals.</td>
<td>A more physiologic test of glucose disposal than i.v. loads.</td>
<td>Measures glucose tolerance and not aligned to insulin sensitivity per se.</td>
</tr>
<tr>
<td><strong>Fasting methods</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fasting insulin</td>
<td>( \beta )-cell insulin secretion is steady and will be lower in those with better SI in the fasting state.</td>
<td>Quick and inexpensive single assay. Easy to develop reference ranges and interpret.</td>
<td>Assumption of fasting and steady-state metabolism. Cannot be used in diabetic patients as a measure of IR due to insulin-based treatment.</td>
</tr>
<tr>
<td>Glucose:insulin ratio</td>
<td>Used by some researchers in women with PCOS. Essentially corresponds to (1/\text{insulin} ) in the normal fasting glucose range.</td>
<td>Quick and inexpensive.</td>
<td>Cannot be used in diabetic patients; would paradoxically have a higher G:I ratio, indicating better SI.</td>
</tr>
</tbody>
</table>
Table 23.1 (continued)

<table>
<thead>
<tr>
<th>Method name</th>
<th>Method summary</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>HOMA-IR</td>
<td>Homeostasis model assessment (HOMA) is a mathematical model of glucose and insulin interactions. HOMA-IR = insulin (U mL(^{-1})) * glucose (mmol L(^{-1}))/22.5. The denominator of 22.5 is a normalizing factor; i.e., a normal insulin*glucose. Therefore a normal HOMA-IR = 1. A “HOMA2” model has also been developed. Useful for SI evaluation in individuals with glucose intolerance, mild to moderate diabetes, etc. Can be used in large studies.</td>
<td>Not suitable in those with impaired or β-cell function. Poorly validated in animal models.</td>
<td></td>
</tr>
<tr>
<td>Quantitative insulin sensitivity check index (QUICKI)</td>
<td>Since fasting insulin levels are skewed, log-normalized data can be used in models. QUICKI = 1/Log (insulin U mL(^{-1})) + Log (glucose, mg dL(^{-1})). Better correlation with SI by EUC than HOMA-IR. Among the best validated measures in humans.</td>
<td>Log-normalized data unattractive for clinical use.</td>
<td></td>
</tr>
</tbody>
</table>

IR, insulin resistance; PCOS, polycystic ovarian syndrome; SI, sensitivity to insulin.

T2DM risk. Interestingly, as recently reviewed [2], these simple characteristics give good prediction of T2DM risk (area under receiver operator curve around 80%) and thus the benchmark for biomarkers to meaningfully improve prediction beyond simple measures is high.

**Concept of ectopic fat as a cause of tissue insulin resistance**

It appears that individuals prone to T2DM (based in part on their nonmodifiable characteristics) show a greater propensity to accumulate visceral or ectopic fat for a given weight. Interestingly this characteristic, in turn, may be a downstream consequence of “impaired” subcutaneous fat storage capacity, the mechanisms of which deserve further research. As an extreme example of this concept, lipodystrophic individuals have an impaired ability to store subcutaneous fat and, as a consequence, they accumulate fat in visceral and ectopic tissues and so have marked insulin resistance [3]. It is also apparent that certain ethnic groups at heightened diabetes risk may have a tendency to store fat centrally sooner (i.e., at lower average BMI) than European whites and consequently develop diabetes at lower average BMI values and around a decade earlier [4]. At the other extreme, there are many individuals, particularly women, who despite attaining very high BMIs, even as high as 50 to 60 kg m\(^{-2}\), remain insulin sensitive and normolipemic. Imaging studies have shown these individuals to have low levels of visceral and ectopic fat but a high subcutaneous fat content [5]. In other words, the location of fat storage (subcutaneous vs. visceral/ectopic) appears critical as to the BMI and time that an individual develops metabolic complications linked to insulin resistance. Figure 23.2 summarizes the concept of ectopic fat and its importance to insulin resistance.

**Pathways linked to insulin resistance**

From the preceding observations on the actions of insulin and the role of ectopic fat in exacerbating insulin resistance, a number of linked metabolic perturbances can be predicted. The most clinically relevant and interesting pathways are discussed later with an attempt to draw out potential causal associations where relevant.

**Lipid alterations in insulin resistance syndrome**

Given insulin’s role in suppressing FFA release, it is clear that insulin resistance must be associated with raised triglyceride levels. Similarly, given that in most individuals triglyceride levels inversely correlate with HDL-cholesterol levels, low HDL-cholesterol is a commonly recognized feature in insulin resistance. Beyond these two changes, it has been clearly shown that higher apolipoprotein B levels and greater preponderance of small dense LDL are linked to insulin resistance [6]. Whilst this lipid pattern is often considered to be linked to increased vascular risk in the prediabetes individual, the causal nature of these associations remains debated. Furthermore, it is now clear that elevated triglyceride levels in most individuals more strongly indicate diabetes than vascular risk [7]. There is also now strong evidence that once cholesterol and HDL-cholesterol are included in cardiovascular risk scores, addition of triglyceride or other lipid fractions adds minimally to prediction of cardiovascular events [8,9]. Despite this, evidence from a large genetic study is potentially consistent with a causal role for triglycerides in heart disease [10]. Thus it is possible for a risk factor to be causal, but offer redundant information in terms of risk scores.

**Liver markers and insulin resistance**

In terms of chronology, gain of liver fat appears to immediately precede development of diabetes in most individuals whereas
The insulin resistance syndrome

Muscle insulin resistance appears to be a longer standing and earlier abnormality [11]. Considerable recent work has focused on the role of the liver in the development of diabetes and, linked to this, the condition of nonalcoholic fatty liver disease (NAFLD) has been shown to be very common (>50%) in T2DM and is linked to hepatic insulin resistance [12]. Several factors released by the liver have been associated with insulin resistance as shown in Table 23.2. These include ALT, GGT, SHBG, IGF-1, Ferritin, and CRP [13–17]. In terms of clinical utility, clinicians now commonly recognize the value of ALT and GGT to help define risk of NAFLD. However, whilst ALT and GGT have been shown to be linked to incident diabetes, they are not, and are unlikely to be, included in diabetes risk scores. SHBG is measured in many centers together with testosterone to derive a free androgen index ([testosterone / SHBG]∗100), a parameter used in the diagnosis of polycystic ovarian syndrome (PCOS), a condition linked to insulin resistance and obesity [18]. In terms of causality, preliminary evidence suggests SHBG may be causally linked to diabetes [19] but, interestingly, not to insulin resistance. Further basic and clinical work in this area would be useful to tease apart the causal pathway.

Fetuin-A (also called alpha-2-HS-glycoprotein) is a liver-derived glycoprotein involved in calcium and phosphate transport. Recent evidence indicates that Fetuin-A may be implicated in insulin sensitivity through inflammatory signaling pathways [20]. Consequently clinical and epidemiologic studies of circulating levels of this biomarker, and its implications for disease, are now emerging.

Adipose-derived markers and insulin resistance

Whilst rising BMI or waist circumference are clearly linked to insulin resistance and risk for T2DM, a number of adipose-derived factors have attracted interest with respect to determining insulin resistance. Adiponectin, as recently reviewed [21], has been the most notable of these biomarkers. Adiponectin is unlike the other adipocyte hormones in that its concentrations decline with rising obesity. Therefore low levels predict higher risk for T2DM but there are a number of complexities in this relationship. First, it remains unclear if adiponectin measurement aids diabetes risk prediction. For instance, in the British Regional Heart Study men with adiponectin concentrations in the top third of the distribution had a 60% (95% CI 30–77%) lower risk of incident T2DM.

Figure 23.2 Simple concept of ectopic fat and development of insulin resistance and frank type 2 diabetes. This figure provides a simple conceptual illustration of the development and location of ectopic fat (and thus worsening insulin resistance) in individuals once they have “overwhelmed” their ability to store safe subcutaneous fat. Certain factors such as gender (females having greater storage capacity), genetics (with family history of diabetes as a broad proxy measure), ethnicity (e.g. South Asians) and aging have relevance to an individual’s ability to store fat subcutaneously. In temporal terms, it may be liver fat accumulation is closer to the time of development of diabetes whereas muscle insulin resistance is an earlier development. Perivascular fat may contribute to vascular dysfunction via a process of adverse vasocrine signaling leading in turn to impaired nutrient blood flow—that is, vascular insulin resistance. Finally, some recent evidence indicates excess fat may also accumulate in the pancreas to contribute to β-cell dysfunction, and thus development of diabetes. IR, insulin resistance.
Table 23.2 Relationship and potential causal role of factors emerging from different pathways with insulin resistance

<table>
<thead>
<tr>
<th>Novel predictor</th>
<th>Background</th>
<th>Examples of association with IR in a variety of clinical studies*</th>
<th>Evidence of causality</th>
<th>Other relevant points</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Liver-derived factors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALT/GGT</td>
<td>Liver transaminases are elevated in hepatic fat accumulation, NAFLD. When AST rise starts to parallel ALT rise, NASH is indicated.</td>
<td>Association with SI during EUC: $\beta = -0.18$ in men and $-0.15$ in women (for ALT), and $\beta = -0.17$ in men and $-0.11$ in women (for GGT) after adjustment for age and center [13].</td>
<td>Markers of ectopic fat rather than causal, though liver fat may be causally associated with insulin resistance.</td>
<td>ALT and GGT strongly positively associated with diabetes risk [64].</td>
</tr>
<tr>
<td>SHBG</td>
<td>Testosterone and estrogen circulate bound to liver-derived SHBG.</td>
<td>Association with SI during EUC: $r = 0.15$ after adjusting for age, sex, and body fat. Null association after further adjusting for hepatic fat [14].</td>
<td>Evidence from MR studies that SHBG may be causal in diabetes, but no association with IR [19].</td>
<td>Monosaccharide-induced lipogenesis reduces hepatic SHBG expression [65].</td>
</tr>
<tr>
<td>Ferritin</td>
<td>Iron transporter. Circulating levels also increase during acute phase response.</td>
<td>Association with HOMA-IR $r = 0.20$ in Korean men and $r = 0.15$ in Korean women. Only associated with HOMA-IR in men after multiple adjustments [15].</td>
<td>No evidence for causality.</td>
<td>Unlikely to be used for diabetes prediction.</td>
</tr>
<tr>
<td>IGF-1</td>
<td>Insulin-like growth factor-1.</td>
<td>Associated with HOMA-IR after adjusting for age and sex ($r = -0.28$) [16].</td>
<td>Evidence in mice that IGF-1 regulates insulin and growth hormone balance in SI [66].</td>
<td>Several analogues available, but none examined for glycemic control.</td>
</tr>
<tr>
<td>CRP</td>
<td>Classical marker of the acute phase response.</td>
<td>Associated with SI (from IV GGT) after adjusting for age, sex, center, ethnicity, and smoking: $r = -0.37$ [61]. Strong positive association with HOMA-IR by quartiles in Framingham Offspring study [7] and other studies [67].</td>
<td>Clear evidence against causality [68], though upstream IL-6 may be more relevant.</td>
<td>Predictive clinical utility of CRP for CVD now also in doubt.</td>
</tr>
<tr>
<td>Fetuin-A</td>
<td>Fetuin-A is a mineral carrier. Deficiency is associated with soft tissue calcification.</td>
<td>Weak but significant association with HOMA-IR in a community study of Chinese participants ($\beta = 0.03$) after multiple adjustments [72].</td>
<td>Fetuin-A may interact with toll-like receptor 4 in altering insulin sensitivity in mouse models [20].</td>
<td>Increased Fetuin-A is associated with slightly reduced CVD risk among those without diabetes [69].</td>
</tr>
<tr>
<td><strong>Adipose-derived factors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adiponectin</td>
<td>Adipose-derived hormone inversely associated with BMI. Putative insulin sensitizer.</td>
<td>Univariable association with SI during EUC $r = 0.28$ in women and $r = 0.42$ in men without DM [64]. After adjusting for age, sex, and measures of adiposity, this association was $r = 0.23$ [70,71].</td>
<td>Strong evidence of causality in animal models. Recent evidence goes against adiponectin as causally related to IR or type2 diabetes in man [24].</td>
<td>High adiponectin associated with lower diabetes but in older individuals or those with CVD, also associated with greater total and CVD mortality.</td>
</tr>
<tr>
<td>Leptin</td>
<td>Adipose-derived hormone involved in appetite regulation. +ve association with BMI.</td>
<td>Univariable association with SI (glucose uptake) during EUC $r = -0.59$ in women and $r = -0.83$ in men without DM [73].</td>
<td>Absolute leptin deficiency causes IR. High levels of leptin in obesity possibly a result of “leptin resistance” causing lack of satiety.</td>
<td>Leptin not clinically useful to enhance diabetes prediction over simpler measures of obesity.</td>
</tr>
<tr>
<td><strong>IL-6</strong></td>
<td>Inflammatory marker—causes hepatic acute phase response. Potential causal role in CVD.</td>
<td>Association with SI during EUC $r = -0.23$ in a study of African Americans [71]; robust to adjustment for fat mass.</td>
<td>Limited evidence an IL-6 blocker increases insulin sensitivity in RA patients [74], plus indirect genetic evidence for potential causal link between IL-6R and diabetes [26].</td>
<td>Although potentially causal, robust intervention studies currently lacking.</td>
</tr>
<tr>
<td><strong>PAI-1</strong></td>
<td>Principal inhibitor of t-PA, thus an inhibitor of fibrinolysis.</td>
<td>Association with SI during EUC $r = -0.21$ in a study of African Americans [71]; robust to adjustment for fat mass. Strong positive association with HOMA-IR by quartiles in Framingham Offspring [17].</td>
<td>PAI-1 deficient mice do not develop obesity or IR on high-fat diet [75,76]. Lack of data in humans.</td>
<td>Tiplaxtinin is a small molecule inhibitor (not suitable for long-term treatment). ACE inhibitors and insulin sensitizers also reduce PAI-1.</td>
</tr>
</tbody>
</table>

### Endothelial-derived factors

| t-PA antigen | Tissue plasminogen activator is involved in fibrinolysis. | Moderate positive association with HOMA-IR in older men ($r = 0.25$ after adjusting for WC). Similar results in both genders [77]. | Elevated levels are a result of endothelial dysfunction; a marker of disease rather than a cause, commensurate with fibrinolytic role. | t-PA can also be released from liver and thus, hard to fully distinguish whether its links to IR represent hepatic or vascular origins. |
| Adhesion molecules | P- and E-selectin, ICAM-1, VCAM-1 are expressed on endothelial wall to allow leukocyte extravasation. Solubilized circulating forms are detectable in plasma. | VCAM-1 has weak positive associations with HOMA-IR in crude analysis and after adjusting for age, sex, and BMI ($r = 0.09$) [78]. In one study of adults, after adjusting for age and sex, only E-selectin was associated with HOMA-IR [79]. | No strong evidence of causality for the circulating isoforms. Reasons for cleavage from endothelial wall unclear. | The role of soluble adhesion molecules in human diseases is not clear. |
| VWF antigen | VWF is released from endothelial wall as a result of vascular dysfunction. Involved in the coagulation cascade. | Moderate positive association with HOMA-IR by quartiles in Framingham Offspring [17]. Weak positive association in older men ($r = 0.10$) after adjusting for WC [80]. | More likely to be a downstream cause of diabetes complications than to cause IR per se. | VWF also only weakly associated with diabetes risk. |

### Nutritional factors

| Vitamin C | Classical antioxidant vitamin, often reported to limit/reverse oxidative stress. | No strong association with HOMA-IR among adolescents [81]. | Mixed evidence from primarily small trials [82]. | Meta-analysis of RCTs indicates strong evidence against antioxidant vitamin supplements protecting from risk of mortality. |
| Vitamin D | “The sunshine vitamin.” Has been related to a wide variety of chronic diseases. | No strong association with SI or HOMA-IR comparing those with high and low 25OHD concentrations after adjusting for confounding variables [83]. | Insufficient evidence [84]. | Large trials of vitamin D supplementation in middle-aged populations are underway. Current meta-analyses cast doubt on benefit of supplementation in chronic disease. |

### Cardiac biomarkers

| ANP/BNP | A-type and B-type natriuretic peptide produced in response to myocardial stretch. | BNP is inversely associated with HOMA-IR in healthy young people [85], although the association is null or positive in older cohorts or those with prevalent disease. | Evidence from MR studies that BNP is causally linked to insulin sensitivity [29]. | BNP release from cardiac tissue may therefore function not only to improve vascular function but also to enhance metabolic efficiency. |

*Note: $r$ or $\beta$ values should not be directly compared between studies given differing adjustment models. EUC, euglycemic-hyperinsulinemic clamp; ISI, insulin sensitivity index; RCT, randomized controlled trial; SI, sensitivity to insulin; WC, waist circumference.*
although adjustment for insulin resistance attenuated the association [22]. Second, in older individuals or those with increased vascular risk, high, not low, adiponectin predicted increased risk of cardiovascular disease and mortality [23]. Risk associations are thus highly context dependent.

With respect to causality, a recent Mendelian randomization study did not support a causal role for adiponectin in insulin resistance and T2DM [24]. Furthermore, other data have shown that low adiponectin in humans is more likely to be a downstream signal of hyperinsulinemia, than a causal upstream determinant of insulin resistance [25].

With respect to other adipokines, leptin has also been widely suggested as a cause of insulin resistance, but in most individuals higher leptin is a marker of percent fat mass. Due to the great degree of correlation with fat mass (i.e., information that can be gained from other simple measures of adiposity) it is unlikely to ever be used in clinical practice to predict diabetes. More interesting adipokines are: (i) IL-6, which is also released from fat cells; there is emerging evidence of a causal association of IL-6R with cardiovascular disease [26] and potentially diabetes, and (ii) PAI-1, an inhibitor of fibrinolysis [27]. There is preliminary evidence of a potential causal association with diabetes. Further work on IL-6 and PAI-1 seem warranted.

Endothelial-derived markers and insulin resistance

There is a wealth of data suggesting a potential role for endothelial dysfunction in insulin resistance. Although the direction of causality remains hotly debated, circulating elevations in some endothelial-derived factors, cell adhesion molecules, in particular E-Selectin, ICAM-1, and t-PA have been shown to predict risk for T2DM, independently of other predictors [2]. However, t-PA can also be released from the liver so its link to insulin resistance may be via differential pathways. VWF has also been linked to insulin resistance but more weakly than other parameters. Finally, microalbuminuria is often considered to reflect vascular dysfunction at the level of the kidney and, interestingly, its levels in urine have long been known to be related to insulin resistance in nondiabetic subjects [28].

Cardiac biomarkers and insulin resistance

There is emerging evidence that as well as metabolism impacting cardiac function, cardiac function may also influence metabolism. Natriuretic peptides are produced by the myocardium in response to mechanical stretch as a downstream result of a range of cardiac and vascular pathologies. A-type and B-type natriuretic peptides exert natriuretic and diuretic effects to limit elevation in blood pressure and fluid accumulation. Circulating levels of BNP and its inactive N-terminal metabolite (NT-proBNP) are used for the clinical exclusion of heart failure. However, evidence from genetic studies shows that elevated BNP protects generally healthy people from diabetes [29]. How do natriuretic peptides protect from insulin resistance? There is evidence that natriuretic peptides cause lipolysis from adipose tissue, and also result in an increase in adiponectin [30]. Thus natriuretic peptides may go some way to “explaining” the adiponectin paradox described earlier; high levels in healthy people are protective, but high levels arising from pathology are adverse signals. Whilst these latter relationships are not commonly discussed in relation to the insulin resistance syndrome, they serve to show that insulin resistance is being linked to some unexpected parameters and in different tissues from those commonly associated with insulin resistance.

Insulin resistance as cardiovascular risk factor?

Given that insulin resistance is associated with dyslipidemia, dysglycemia, high blood pressure, inflammation, thrombotic pathways, and endothelial dysfunction, many have suggested that insulin resistance must be a strong cardiovascular risk factor. Yet, the evidence base for this assumption is relatively limited and there are no large studies linking gold standard clamp-based insulin resistance measurements to subsequent cardiovascular outcomes. The multinational Relationship between Insulin Sensitivity and Cardiovascular disease (RISC) study recently reported weak cross-sectional associations of insulin sensitivity (measured by clamp) with carotid intima media thickness in men but not in women, and insulin sensitivity was not associated with three-year changes in cIMT [31]. To address this important issue, we recently related circulating concentrations of fasting and nonfasting insulin, as well as pro-insulin, to incident coronary heart disease in a meta-analysis of 19 Western prospective studies [32]. The findings showed that in a comparison of individuals who had circulating levels of each of these markers in the top third with those in the bottom third of the population, the odds ratio for CHD was 1.12 (95% CI 0.98–1.28) for raised fasting insulin, 1.35 (1.14–1.60) for raised nonfasting insulin, and 2.23 (1.65–3.00) for raised pro-insulin. These results also suggest that links between CHD risk and fasting or nonfasting insulin levels are likely to be more modest than previously suspected. Powerful genetic data consortia are required to address the causal role of insulin in CVD with confidence. Clearly, this area requires further research but, currently, the best available data do not support hyperinsulinemia as a strong risk factor for vascular disease in people without diabetes.

Other markers and future research

The foregoing observations have detailed the best established risk factors associated with insulin resistance (Figure 23.3) as well some novel emerging risk factors as additionally detailed in Table 23.2, including nutritional measures such as Vitamin D. These are not comprehensive but rather a selection of emerging pathways of interest. Clearly, future research should increasingly harness genetics as well as observational data to better determine causal risk pathways. Similarly, further data are needed to better determine the extent to which insulin resistance is, or is not, linked to cardiovascular risk.
Introducing insulin resistance to the clinic: the metabolic syndrome

The second section of this chapter now describes how researchers and clinicians have taken commonly measured markers associated with insulin resistance and created risk algorithms which were subsequently tested for clinical utility.

The cluster of commonly measured risk factors for cardiovascular disease and T2DM, which are associated with insulin resistance and are recognized to commonly coexist, have become known as the “metabolic syndrome.” The concept of the metabolic syndrome was initially proposed by Reaven who suggested a conceptual framework to link these apparently unrelated clinical and biochemical characteristics [33].

What are the metabolic syndrome criteria?

Features of metabolic syndrome include hypertension, dyslipidemia (specifically hypertriglyceridemia with low HDL-cholesterol), dysglycemia, and central obesity though it should be highlighted that the pathophysiologic basis for the association between these factors remains unidentified. Different national and international organizations have produced various overlapping but different iterations of diagnostic criteria for metabolic syndrome over the last 15 years. Early criteria from the World Health Organization (WHO), International Diabetes Federation (IDF) and National Cholesterol Education Panel (NCEP) for diagnosing metabolic syndrome are listed in Table 23.3 [34–36]. Common to all was the requirement to fulfill three out of five criteria but not all stipulated an essential criterion. For example, WHO metabolic syndrome [36] criteria required the presence of some biochemical evidence of insulin resistance or dysglycemia, to which other characteristics were then added. By contrast, IDF criteria [35] required evidence of central obesity while NCEP ATP III criteria made no such stipulation. Therefore, WHO criteria were most directly linked with Reaven’s proposed, but still unexplained, observation linking insulin resistance with hypertension and dyslipidemia as an explanation for the excess cardiovascular risk in diabetic patients. One immediate practical difficulty was the fact that fasting insulin levels were seldom measured in a routine clinical setting and that a definitive insulin resistance diagnosis was poorly defined. Therefore the change in focus of the IDF criteria to central obesity rendered these criteria not only easier to assess, but also raised awareness of obesity as being a key player in the majority at risk of developing T2DM. The IDF and NCEP criteria were also very easily applied to many existing large epidemiologic studies, leading to the publication of huge swathes of epidemiologic research of variable quality over the last two decades exploring the clinical utility of metabolic syndrome. Importantly, the availability of various criteria for identifying metabolic syndrome and the knowledge that the relationship between risk of cardiometabolic disease and central...
Table 23.3 World Health Organization (WHO), International Diabetes Federation (IDF) and National Cholesterol Education Panel (NCEP) criteria for diagnosing metabolic syndrome

<table>
<thead>
<tr>
<th>WHO (modified) criteria [36]</th>
<th>IDF criteria [35]</th>
<th>NCEP criteria [34]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Essential criterion</td>
<td>Plus at least two of:</td>
<td>Plus at least two of:</td>
</tr>
<tr>
<td>Fasting hyperinsulinemia</td>
<td>BMI &gt; 30 kg m⁻² or waist:hip ratio &gt; 0.90</td>
<td>Men &gt; 102 cm or women &gt; 88 cm waist circumference</td>
</tr>
<tr>
<td>(highest 25% of nondiabetic population), or Type 2 diabetes mellitus, or Impaired glucose tolerance, or Fasting plasma glucose ≥ 6.1 mmol L⁻¹</td>
<td>≥ 5.6 mmol L⁻¹ or previously diagnosed type 2 diabetes</td>
<td>≥ 6.1 mmol L⁻¹</td>
</tr>
<tr>
<td>Central obesity</td>
<td>≥ 140/90 mmHg or on therapy</td>
<td>≥ 130/85 mmHg or on therapy</td>
</tr>
<tr>
<td>Fasting plasma glucose (fasting)</td>
<td>≥ 2.0 mmol L⁻¹</td>
<td>≥ 1.7 mmol L⁻¹ or receiving specific therapy</td>
</tr>
<tr>
<td>Blood pressure</td>
<td>Reduced HDL-cholesterol</td>
<td>≥ 1.0 mmol L⁻¹</td>
</tr>
<tr>
<td>Triglycerides (fasting)</td>
<td>Systolic ≥ 130 mmHg and/or diastolic ≥ 85 mmHg or drug therapy with a history of hypertension</td>
<td></td>
</tr>
<tr>
<td>HDL-cholesterol</td>
<td>Elevated blood pressure</td>
<td>≥ 5.6 mmol L⁻¹ or glucose-lowering therapy</td>
</tr>
<tr>
<td>Other</td>
<td>Elevated fasting plasma glucose</td>
<td>–</td>
</tr>
<tr>
<td>Microalbuminuria</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

Obesity varies according to ethnic group led to confusion in clinical practice and controversy in published literature, though controversy has not been limited to this specific issue. To address recognized shortcomings, a harmonized set of criteria was produced by a working group on behalf of the International Diabetes Federation Task Force on Epidemiology and Prevention, the National Heart, Lung, and Blood Institute, the American Heart Association, the World Heart Federation, the International Atherosclerosis Society, and the International Association for the Study of Obesity in 2009 [37] (see Table 23.4). As in previous criteria, three out of five criteria are now required to be identified as having metabolic syndrome. However, the presence of central obesity as an essential criterion was removed with the result that the presence of any three criteria now constitutes a diagnosis of metabolic syndrome. Furthermore, flexibility has been introduced into the definition of metabolic syndrome as population- and country-specific definitions for elevated waist circumference are required, reflecting the understandable lack of agreement on how best to record and define central obesity. Suggested current thresholds for elevated waist circumference, as recommended by various organizations, are provided in the document (see also Table 23.5). The authors propose that further research is required to properly delineate the relationship between waist circumference and cardiometabolic risk in all ethnic groups and suggest that particular health systems may even base the choice of waist circumference threshold on local considerations, including healthcare funding.

The purpose of the metabolic syndrome concept

The purpose behind introduction of the metabolic syndrome into clinical practice and published epidemiologic research appears to have been the desire to facilitate identification of those at risk of experiencing a cardiovascular event and/or developing T2DM in the medium to long term using a simple tool. Rationale for using metabolic syndrome has been summarized as: (i) facilitating research into a possible unifying pathophysiologic explanation for these strongly associated
risk factors, (ii) quantifying chronic disease risk and allowing between-country comparison, (iii) guiding clinical management decisions, and (iv) providing an easily understood public health message [38]. Those involved in developing and promoting the concept of the metabolic syndrome have described it as a potentially valuable global public health tool with the ability to identify high-risk individuals at a young age. Undoubtedly, one positive effect of the adoption of the term has been to familiarize healthcare professionals with the potentially deleterious effects of obesity and to provide an easily understood term, or clinical label, to describe a commonly encountered type of patient. Furthermore, appreciation of those with metabolic syndrome has likely raised the profile of lifestyle advice and increased the provision of advice to those at risk of developing diabetes.

The clinical utility of metabolic syndrome has been controversial, however, with proponents arguing that identifying a patient with metabolic syndrome is a simple and effective method to identify someone at potentially long-term risk of cardiometabolic risk [39] but with skeptics arguing that the label of “metabolic syndrome” adds little to established methods for risk calculation for cardiovascular disease and T2DM [40 – 42]. Some key topics of debate have been the presence of various differing sets of criteria, now addressed by the introduction of the harmonized criteria already referred to, and the fundamental question of the role and value of metabolic syndrome in clinical practice. It is this latter point that remains unresolved. Some of the more humorous and incisive pieces in the medical literature in recent years have addressed this matter [43 – 45].

Since the introduction of the Framingham Risk Score (based on classical cardiovascular risk factors namely hypercholesterolemia, hypertension, smoking), supplemented more recently with ever better performing risk scores in the United Kingdom such as QRISK2 [46] and ASSIGN [47] which incorporate additional risk factors like family history and social deprivation, clinicians have had at their disposal effective tools for prediction of cardiovascular events thereby facilitating allocation of effective risk-lowering therapies to those at highest risk. The presence of metabolic syndrome is itself unequivocally associated with increased risk of cardiovascular disease, as confirmed in various studies [48 – 50], but this is unsurprising given that it includes some powerful cardiovascular risk factors, namely hypertension and low HDL-cholesterol. In the view of many, the key question has therefore been whether identifying a patient as having metabolic syndrome yields additional predictive information on top of these risk scores. In other words, to be of clinical value, any new risk predictor must be more effective than current risk predictors to demonstrate clinical utility. It has now been well demonstrated that identification of metabolic syndrome cannot compete with the cardiovascular predictive capabilities of established cardiovascular risk scores and also that it adds minimal incremental information [51 – 53]. This apparent under-performance is unavoidable given the necessary absence of three of the strongest known cardiovascular risk factors namely age, low-density lipoprotein (LDL) cholesterol, and smoking status. To illustrate this problem with a clinical example, two theoretical but commonly encountered patients are listed in Table 23.6 (as previously described [42]).
Although both are hypertensive, only Mr. A fulfills all criteria for metabolic syndrome (5/5), unlike Mr. B. Nonetheless, Mr. B is at considerably greater risk of vascular disease due to his older age, his higher LDL-cholesterol, and his smoking status. In addition, individual components of the metabolic syndrome often outperform the overall metabolic syndrome unlike established cardiovascular risk scores [54–56].

Where metabolic syndrome arguably makes a stronger case for routine clinical use is in the prediction of developing T2DM, as demonstrated in previous epidemiologic studies [49,57] and in keeping with Reaven's proposed unifying hypothesis, insulin resistance. Three of the five variables in the metabolic syndrome criteria are more strongly linked to diabetes than cardiovascular disease, namely elevated triglycerides, central obesity, and dysglycemia. As an example, the Emerging Risk Factors Collaboration’s (ERFC) comprehensive analysis established that impaired fasting glycemia is only weakly associated with incident cardiovascular events [58]. By contrast, IFG’s association is an order of magnitude more powerful with risk of developing diabetes than with risk of cardiovascular disease [59]. The cases of BMI and hypertriglyceridemia are analogous. Regardless, knowledge of an individual’s fasting blood glucose provides at least the same predictive information as knowledge of metabolic syndrome status for identifying those most likely to develop T2DM.

### Statistical considerations and other issues

When designing any risk scoring system, those proposing such a score must weigh up various competing factors. Metabolic syndrome benefits from the inclusion of clear (and now unambiguous, with the exception of waist circumference) criteria. However equal weighting is given to the five criteria while it is known that they do not contribute equally to the development of cardiometabolic disease. As noted above, two of the five criteria are strongly linked to cardiovascular disease but only weakly to risk of developing diabetes. The converse is true of the other three variables. Secondly, consideration of any biomarker as a continuous variable adds predictive power while dichotomization of data limits power substantially. One component, serum triglycerides, has a moderately high within-subject biologic variability of 20% with the result that classification of individuals with triglyceride levels close to 1.7 mmol L⁻¹ may give different results from day to day [60]. Other potential weaknesses include the reliance on fasting samples while there is a move towards using nonfasting sampling for both diabetes (HbA1c) and cardiovascular (lipids, HbA1c) risk estimation and diagnosis. Some also challenge the inclusion of those with established T2DM and/or cardiovascular disease in the metabolic syndrome based on the assertion that it adds little to their clinical management [38,40]. Finally, metabolic syndrome describes relative risk of cardiometabolic disease and not absolute risk.

### Treatment of patients with features of the metabolic syndrome

The prevalence of individuals with the metabolic syndrome or components thereof is undeniably high and constantly increasing. Despite control of major cardiovascular risk factors such as LDL-cholesterol, blood pressure, and smoking, a combination of factors regularly leads to a failure to address obesity or improve glycemic control with a resultant residual higher cardiovascular risk than is observed in the normal population. Therefore, attention must be focused on what additional approaches can be brought to bear to reduce this residual risk. While an in-depth discussion of outcome trial data regarding drugs aimed at lowering triglycerides or increasing HDL-cholesterol is beyond the scope of this chapter, their results can be summarized as disappointing thus far. Whether we should target the hypothesized upstream insulin resistance, and if so with what tools, remains unproven for reducing cardiovascular risk. Those with impaired fasting glycemia and impaired glucose tolerance are not routinely prescribed glucose-lowering agents although cardiovascular endpoint trials of relevant therapies such as metformin are either underway or infeasibility studies. Lifestyle modification is commonly recommended and has been shown to reduce the development of T2DM in those at risk [61,62], but data supporting a reduction in cardiovascular outcomes remains elusive. Problematically, while intensive lifestyle advice can be provided in the context of a clinical trial, significant improvement remains difficult to achieve and sustain outside the research environment. Large observational studies of weight loss surgery strongly suggest a reduction in risk of developing diabetes [63] but access to such procedures is necessarily limited to a subset of the population based on availability of resources.

### Current thinking regarding metabolic syndrome

In November 2008, a WHO Expert Consultation was undertaken to evaluate the utility of the concept of metabolic syndrome.

<table>
<thead>
<tr>
<th>Mr. A</th>
<th>Mr. B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Waist circumference (cm)</td>
<td>105</td>
</tr>
<tr>
<td>Triglyceride (mmol L⁻¹)</td>
<td>2.2</td>
</tr>
<tr>
<td>HDL-cholesterol (mmol L⁻¹)</td>
<td>0.9</td>
</tr>
<tr>
<td>Fasting plasma glucose (mmol L⁻¹)</td>
<td>6.2</td>
</tr>
<tr>
<td>Blood pressure (mmHg)</td>
<td>145/90</td>
</tr>
<tr>
<td>Age (years)</td>
<td>40</td>
</tr>
<tr>
<td>Smoking status</td>
<td>No</td>
</tr>
<tr>
<td>LDL-cholesterol (mmol L⁻¹)</td>
<td>3.0</td>
</tr>
<tr>
<td>Metabolic syndrome</td>
<td>Yes (5/5 criteria)</td>
</tr>
<tr>
<td>Framingham risk score (%)</td>
<td>8</td>
</tr>
</tbody>
</table>

Table 23.6 Hypothetical example of cardiovascular risk in two individuals, one with and one without metabolic syndrome.
syndrome in four areas: pathophysiology, epidemiology, clinical work, and public health [38]. Much of the ground addressed in the report of the consultation is covered in the earlier sections of this chapter. The consultation’s conclusions are ones which the authors of this chapter are in agreement with. They made the simple but important observation that while metabolic syndrome has gained traction as a widely recognized concept, a formal diagnosis of metabolic syndrome is seldom made and the condition is rarely included in guidelines for the treatment and prevention of cardiometabolic disease. The group rightly called for a halt to studies comparing different versions of the metabolic syndrome and also emphasized that research should be focused on identifying the mechanism or mechanisms underlying the clustering of metabolic risk factors, rather than attempting to further refine diagnostic criteria. The main conclusion was therefore that while metabolic syndrome represents a useful educational concept, it has limited practical utility as a diagnostic or management tool.

Conclusion

The terms “insulin resistance syndrome” or “metabolic syndrome” are widely used in clinical research but their precise definitions remain weakly defined and their clinical utility questionable. The present chapter has summarized numerous biomarkers associated with insulin resistance and attempted, where possible, to determine which may be causally linked to insulin resistance, and the directionality of such links. A minority of insulin resistance-associated risk factor perturbations (e.g., those related to NAFLD or PCOS) appear relevant to clinical practice whereas simple risk factors (age, gender, obesity, family history, and ethnic origin) predict type 2 diabetes well. Novel blood-based insulin resistance biomarkers seem unable to meaningfully (or cost-effectively) improve diabetes prediction.

In terms of vascular disease, and contrary to popular belief, the current data do not support a strong association of insulin resistance (in the absence of diabetes) with cardiovascular disease and, similarly, metabolic syndrome criteria do not improve prediction beyond simpler nonfasting cardiovascular risk scores.

References


**CHAPTER 24**

**β-Cell mass and function in human type 2 diabetes**

Piero Marchetti1 and Ele Ferrannini2

1Department of Endocrinology, University of Pisa, Pisa, Italy
2Department of Clinical and Experimental Medicine, University of Pisa, Pisa, Italy

**Key points**

- β-Cell mass is reduced by 30–40% in type 2 diabetes, with a wide overlap with nondiabetic subjects.
- Both enhanced apoptosis and abnormal autophagy account for β-cell demise.
- Relative excess of α cells, architectural disarray, and reduced connectivity are additional features of diabetic islets.
- Cultured human β-cells from diabetic patients respond to glucose poorly as do β-cells from nondiabetic donors exposed to in vitro hyperglycemia.
- In vivo, impaired β-cell glucose sensing and incretin-mediated secretory potentiation contribute to dysglycemia quantitatively.
- Functional incompetence dominates over β-cell loss in the pathogenesis of diabetes and prediabetes.
- A variable but clinically significant proportion of β-cell dysfunction is reversible.

**Introduction**

Insulin-producing β-cells are the predominant endocrine cell type in pancreatic islets, comprising 50–80% of islet cells [1–3]. Studies with autoptic samples, organ donor specimens, and surgical cases have found that β-cell mass in the human pancreas may vary from 0.6 to 2.1 g, with β-cell volume and area (relative to the pancreatic tissue) ranging from 1.1–2.6% and 0.6–1.6%, respectively [4–15]. Islet insulin content has been reported to be 100–500 μU per islet [16–20], but some islets may contain more than 1000 μU insulin [20]. As islet number in adult subjects also varies widely (0.5–4 million per pancreas) [21], pancreatic insulin content, as measured in a few studies, shows a marked intersubject variability (up to 10-fold) [7,17].

Functionally, β-cells are the most sophisticated of endocrine cells. Insulin must be supplied to body tissues in amounts, time-dynamics and flexibility able to maintain plasma glucose within a narrow concentration interval on a minute-by-minute basis [22–25]. In fact, insulin secretion must cope with acute (meals, exercise, stressful events) and long-term (body weight gain or loss, aging, pregnancy) changes. Whereas the amount of insulin released daily has been reported to range ~20–70 U (mainly depending on body weight) [26,27], capturing the multifaceted features of insulin secretion in vivo is a difficult task [28–30]. However, the recent use of human islets isolated from the pancreas of organ donors has allowed the direct study of β-cell insulin release under different conditions, showing that several of the insulin secretion properties observed in vivo (dynamics, oscillatory pattern, metabolic and/or pharmacologic perturbations) can be reproduced ex vivo [18,31–37]. In this chapter, we will summarize available knowledge on the quantitative and functional differences between nondiabetic and T2DM β-cells, showing that β-cell loss and changes in β-cell secretory phenotype could have a different impact on the onset and progression of type 2 diabetes (T2DM).

**β-Cell mass in type 2 diabetes**

Although quantification of β-cells in the human pancreas is a difficult task, morphometric analyses have been performed by several authors to assess β-cell mass (when the weight of pancreas specimens was available), volume (usually assuming that the islets are spherical), and/or area (insulin-positive proportions in the islets or the pancreatic tissue) [4–7,9,10,12–15,38–46]. The amount of β-cells has been usually (although not always) reported to be significantly reduced in the pancreata of patients with T2DM; Table 24.1 summarizes most of the available data. Early work reported that total islet number was 30–50% lower in histology samples from type 2 diabetic subjects as compared to nondiabetic individuals [38,39]. A reduction of total islet volume in diabetic versus nondiabetic pancreata was also shown [40,41], which became...
more marked when accounting for the presence of amyloid [40]. In the following years, although a few authors were not able to find major differences in β-cell amount in diabetic compared to nondiabetic pancreas specimens [42–44], several studies consistently showed that β-cell quantity is reduced in T2DM. Clark and colleagues reported a 24% reduction in β-cell area in their series of type 2 diabetic pancreas samples [45], and islet β-cell volume and total β-cell mass were found to be significantly lower (~30%) in specimens from Japanese type 2 diabetic patients than in samples from nondiabetic individuals [4]. Similarly, a study conducted in Korea with pancreatic samples from organ donors or following surgical removal found that β-cell volume was approximately 25% reduced in T2DM [5]. Pancreatic autopsic samples from type 2 diabetic patients, subjects with impaired fasting glycemia, and nondiabetic individuals have also been studied (the groups were subdivided into lean or obese based on BMI) [6]. Obese subjects with IFG or diabetes had a 40–60% reduction in β-cell volume in comparison to BMI-matched, nondiabetic cases. In the nonobese group, diabetes was associated with 41% reduction of β-cell volume. In a more recent study [7], the authors analyzed autopsic samples from 57 type 2 diabetic and 52 nondiabetic European subjects, confirming that β-cell mass is lower (35% on average) in the diabetic cases (Figure 24.1).

Similarly, when a comparison was made of pancreata obtained from lean and obese (BMI > 27 kg/m²) diabetic and nondiabetic cadaveric donors [9], it was found that β-cell volume and mass were reduced by approximately 35% in the obese type 2 diabetic cases; however, β-cell amount in lean donors was unaffected by the presence of diabetes [9]. Additional information from later studies indicates that β-cell loss in T2DM may preferentially involve larger islets (in particular when located in the head of the pancreas) [10,13], confirms the potential role of amyloid deposition [46] and the presence of β-cell reduction in prediabetes [14], and suggests that the use of incretins may be associated with changes in β-cell amount [12].

### Table 24.1

<table>
<thead>
<tr>
<th>Authors</th>
<th>Journal and year (reference in this chapter)</th>
<th>Main results</th>
</tr>
</thead>
<tbody>
<tr>
<td>MacLean N &amp; Ogilvie RG</td>
<td><em>Diabetes</em>, 1955 [38]</td>
<td>Reduced islet number</td>
</tr>
<tr>
<td>Westermark P &amp; Wilander E</td>
<td><em>Diabetologia</em>, 1978 [40]</td>
<td>Reduced islet volume</td>
</tr>
</tbody>
</table>
Interestingly, however, a study carried out with the combined use of light and electron microscopy [15] suggested that the loss of β-cells in T2DM as assessed by insulin staining may be overestimated because of degranulation. In summary, the amount of β-cells appears to be reduced in human T2DM, especially in patients with long-standing disease [7]; further work is needed to more precisely quantify the prevalence, time-course and circumstances of β-cell loss.

The role of β-cell death and regeneration

The loss of β-cells in T2DM has been mainly attributed to increased β-cell death due to apoptosis and other forms of cell death [6,9,13,46–49], possibly driven by adverse environmental conditions [50–52] and probably mediated by several intracellular mechanisms [50–54]. Apoptosis is a type of programmed cell death morphologically characterized by cell rounding up, bleb formation and chromatin condensation (Figures 24.2(a) and 24.2(b)). As a matter of fact, in autopic pancreatic samples apoptosis has been shown to be significantly increased in both obese and lean type 2 diabetic cases as compared to BMI-matched, nondiabetic controls [6]. In another study, β-cell apoptosis was found to be threefold increased in obese type 2 diabetic patients [9], although the diabetic condition did not affect apoptosis in lean individuals. Increased β-cell apoptosis in diabetic islets has been confirmed by electron microscopy [47]. In addition, by assessing cytoplasmic histone-associated DNA fragments, it has been observed that islet cell death is significantly enhanced in isolated diabetic islets [18]. These changes are accompanied by increased numbers of cells positive for activated caspase-3 [48] as well as greater activity of caspase-3 and caspase-8 [18], which are key molecules in the induction and execution of apoptosis.

However, forms of programmed cell death other than apoptosis have been described [49]. One involves autophagy, which is a type of cell death that occurs without marked chromatin condensation and is accompanied by massive vacuolization of the cytoplasm [55–57]. In general, normally functioning autophagy has a beneficial role for cells (including the β-cell [58–60]) as it regulates the turnover of aged proteins and eliminates damaged structures and organelles [60]. However, cells that undergo altered autophagy may die in a nonapoptotic manner [55–57]. The presence of autophagy was investigated in β-cells from T2DM and matched nondiabetic subjects [49]. On electron microscopy, there were significantly more dead β-cells in diabetic than control samples; while several of these cells had morphologic signs of apoptosis, massive vacuole overload (suggesting altered autophagy) was associated with a proportion of dead β-cells without signs of apoptosis (Figure 24.2(c)). This proportion was significantly higher in type 2 diabetic samples. It can be therefore concluded that β-cell death is increased in type 2 diabetic patients due to enhanced apoptosis and other forms of cell death, which may contribute to β-cell failure in this disease.

Whether defects of β-cell regeneration also plays a role in the reduction of β-cells in human T2DM is still unclear. β-cell regeneration may essentially occur by replication (proliferation) of existing cells, neogenesis from precursors or transdifferentiation of existing mature cells [61–67]. Normally, insulin-positive cells appear in the human pancreas at around the 8th week of gestation; at 10 weeks postconception all clusters containing more than 10 insulin-positive cells have developed a close relationship with vascular structures [68]. After an additional 2–3 weeks, human fetal islets contain cells independently immunoreactive for insulin, glucagon, somatostatin and pancreatic polypeptide [68]. During prenatal pancreas development, there is a linear increase in fractional β-cell area, which reaches a value of ∼3% at birth [69]. β-cell proliferation is very efficient during fetal life, involving more than 3% of β-cells during week 17–32 of gestation [69,70]; this proportion then tends to decrease and approximates 1.5–2.0% at birth [69,70]. After birth, as shown by the study of autopic pancreatic samples from subjects aged 2 to 20 years, β-cell mass expands several-fold from infancy to adulthood, mainly due to growth in islet size rather than number, a process driven by the rate of β-cell replication [21]. After the age of 20–30 years, however, the rate of islet β-cell replication seems negligible, as shown in two independent studies analyzing lipofuscin accumulation and thymidine incorporation in human β-cells [71,72]. Both these studies concluded that, in human islets, the

![Figure 24.2](image-url)
β-cell complement is fully established by the age of 20–30 years. This implies that β-cell regeneration in adults may occur at a very low rate, if at all.

Nevertheless, β-cell mass in adult human individuals can increase, such as occurs in obesity and during pregnancy [5–7,9,73,74]. In one of the studies mentioned earlier [6], it was observed that nondiabetic obese subjects (with an average BMI of ~35 kg m⁻²) show a 50% increase in β-cell volume as compared to nondiabetic lean subjects (average BMI of ~23 kg m⁻²). Apoptosis and replication did not differ significantly between obese and nonobese cases, and the increase of β-cells was attributed to enhanced neogenesis, as indirectly suggested by the greater number of insulin positive cells in or close to the ducts [6]. However, the same authors, in a more focused article, while confirming the augmented β-cell mass in autopic pancreatic samples of obese subjects, were unable to find differences in terms of apoptosis or regeneration, as assessed by some of the currently available surrogate markers (Ki67, insulin-positive cells in the duct wall) [73]. In another report [9], obesity (average BMI of 31 kg m⁻²) was associated with a twofold increase in β-cell volume compared to lean (average BMI of 24 kg m⁻²) nondiabetic individuals. In this case, the authors found that both β-cell replication and neogenesis were significantly enhanced in the obese samples in the face of similar rates of apoptosis.

Pregnancy is also associated with greater β-cell mass [75]. Although, for obvious reasons, data in humans are very scanty, one study has reported on the morphometry of the pancreatic islets during gestation in humans [74]. The authors collected pancreases obtained at autopsy from women who had died while pregnant, and found that the pancreatic fractional β-cell area was increased by approximately 1.4-fold compared to nonpregnant women [74]. Mean β-cell size was not different, and in pregnancy there were more small islets rather than an increase in islet size or β-cells per islet. No increase in β-cell replication or change in β-cell apoptosis was detected, but duct cells positive for insulin and scattered β-cells were increased with pregnancy, again suggesting, although indirectly, a possible preeminent role of β-cell neogenesis. Interestingly, in a case of gestational diabetes it has been observed that total insulin area was reduced, not due to increased apoptosis but reduced regeneration [76].

The issue of β-cell regeneration has been investigated in a few morphometric studies with pancreases from nondiabetic and type 2 diabetic individuals (Table 24.2). It was initially reported that there was no significant difference in the frequency of β-cell replication (as assessed by Ki67 protein staining) between obese nondiabetic and type 2 diabetic subjects or lean nondiabetic and type 2 diabetic subjects [6]. Similarly, when neogenesis was indirectly quantified by counting duct cells immunoreactive for insulin, no difference was found between the obese or lean nondiabetic and type 2 diabetic cases [6]. However, in another report slightly different results were obtained [9]. When β-cell replication was examined by proliferating cell nuclear antigen (PCNA) staining, this was lower in obese type 2 diabetic than in weight-matched nondiabetic cases; on the other hand, a “neogenic” index calculated from the amount of insulin positive cells or small clusters in the duct walls and/or the acinar tissue was found to be similarly increased in nondiabetic obese individuals and in obese or lean type 2 diabetic patients as compared to nondiabetic lean subjects [9]. Three additional studies were published thereafter [12,14,46], all showing no difference between diabetic and nondiabetic samples in the number of cells co-stained for insulin and Ki67; however, in one report the prevalence of insulin-positive cells in the duct walls was found to be significantly higher in subjects with impaired glucose tolerance and significantly lower in those with T2DM, compared to nondiabetic cases [14].

Altogether, the available information, obtained with the use of surrogate markers and indexes of regeneration, indicates

---

**Table 24.2** Summary of data on β-cell regeneration (type 2 diabetic or “prediabetic” vs. nondiabetic cases) as available from articles retrievable from PubMed (www.ncbi.nlm.nih.gov/pubmed, accessed July 31, 2014)

<table>
<thead>
<tr>
<th>Authors</th>
<th>Journal and year (reference in this chapter)</th>
<th>Main results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Butler AE, et al.</td>
<td><em>Diabetes</em>, 2003 [6]</td>
<td>No difference in the number of cells double positive for insulin and Ki67 between nondiabetic subjects (ND) and individuals with type 2 diabetes (T2DM); no difference in the proportion of insulin-positive cells in the duct walls between ND and T2DM</td>
</tr>
<tr>
<td>Hanley SC, et al.</td>
<td><em>Endocrinology</em>, 2010 [9]</td>
<td>Number of cells double positive for insulin and Ki67 significantly lower in obese T2DM vs. obese ND; “neogenic” index (based on insulin-positive cells/clusters in the duct walls or the acinar tissue) significantly higher in lean T2DM vs. lean ND</td>
</tr>
<tr>
<td>Butler AE, et al.</td>
<td><em>Diabetes</em>, 2013 [12]</td>
<td>No difference in the number of cells double positive for insulin and Ki67 between ND and T2DM; no difference in the proportion of insulin-positive cells in the duct walls between ND and T2DM</td>
</tr>
<tr>
<td>Yoneda S, et al.</td>
<td><em>JCEM</em>, 2013 [14]</td>
<td>No difference in the number of cells double positive for insulin and Ki67 between ND and individuals with IGT or T2DM; prevalence of insulin-positive cells in the duct walls significantly higher in subjects with IGT and significantly lower in those with T2DM, compared to ND</td>
</tr>
</tbody>
</table>
that β-cell replication does not seem to differ between type 2 diabetic and nondiabetic β-cells, whereas neogenesis may or may not be reduced in T2DM. Clearly, even if present, in T2DM β-cell regeneration is not sufficient to replace the β-cells that have been demised.

**The dominance of β-cell functional impairment**

As discussed earlier, the available evidence indicates that an average 30% loss of β-cells is present in the islets of patients with T2DM. However, based on a series of considerations it is unlikely that this β-cell deficit alone is the cause of most cases of diabetes. First, *in vivo* assessment of β-cell function consistently shows a >50% reduction in patients with overt T2DM as compared to nondiabetic subjects, a difference that is amplified when using intravenous glucose [27,77,78] (Figure 24.3). Thus, when β-cell function is indexed as β-cell glucose sensitivity (i.e., the slope of the insulin secretion/plasma glucose dose–response [79], type 2 diabetic patients show a marked inability to respond to changing glucose levels with a prompt rise in insulin release; this deficiency is accentuated in the presence of severe fasting hyperglycemia [80] (Figure 24.4). Of note, while β-cell mass estimates in type 2 diabetic patients overlap with those of nondiabetic subjects widely, the distribution of β-cell functional indices is well separated between type 2 patients and controls (Figure 24.5). It is therefore quantitatively incongruous that a deficit in β-cell mass can be the sole cause of the functional impairment. Secondly, in pancreatectomized patients and in subjects who undergo partial pancreas removal for the purpose of living organ donation, diabetes tends to develop more frequently over the years when β-cell mass declines by more than 50% [81–84]. Thirdly, bariatric surgery is followed by long-lasting resolution of diabetes in a high proportion of type 2 diabetic subjects, in some cases with full recovery of β-cell functional properties [85–87] (Figure 24.6). Finally, it has been clearly shown that any treatment resulting in amelioration of glycemic control is associated with partial reversal of β-cell dysfunction [88–90]. All this is in line with the findings of detailed morphometric studies, showing a very large variability in islet mass, with a wide overlap between nondiabetic and type 2 diabetic individuals [7] (Figure 24.1). Moreover, studies done in parallel with light and electron microscopy have shown that a proportion of β-cells might not be evidenced by insulin immunostaining, due to low content of insulin granules, indicating the presence of living, but functionally incompetent

![Figure 24.4](image-url) Plots of insulin secretion rates against concomitant plasma glucose levels in subjects with normal glucose tolerance (NGT), impaired glucose tolerance (IGT), and type 2 diabetes (T2D, by quartile of fasting hyperglycemia). The mean slope of the fitting functions measures β-cell glucose sensitivity. Source: Redrawn from Ferrannini 2005 [80].

**Oral vs intravenous glucose**

![Figure 24.3](image-url) Insulin secretory response to oral (OGTT) and intravenous glucose (IVGTT) in subjects with normal glucose tolerance (NGT) or type 2 diabetes (T2D). Redrawn from [78]. Note the flat insulin response to intravenous but not oral glucose in T2D.
β-cells [15]. As a consequence, β-cell secretory dysfunction is the dominant defect in the majority of diabetic patients.

This point can also be addressed by studying isolated human islets, which under certain circumstances can be prepared from the pancreas of organ donors (Figure 24.7) [91–93]. Although initial work was not able to show differences in insulin secretion between nondiabetic and type 2 diabetic islets [94,95], studies over the last 2–3 decades have consistently demonstrated defects of insulin release from type 2 human diabetes islets [18,36,96–98]. In a report published in 1994, it was shown that the release of insulin evoked by glucose was lower in type 2 diabetic than in nondiabetic islets [96]. However, the secretory response to a combination of L-leucine and L-glutamine appeared less severely altered [96]. In a more recent study, islets isolated from eight diabetic and nine nondiabetic donors were evaluated by in vitro islet perfusion experiments [98]. Basal insulin secretion was similar between normal and diabetic islets. However, the islets from diabetic donors released less total insulin in response to increasing glucose levels and also exhibited a higher threshold for the initiation of insulin secretion (Figure 24.8). In addition, an equivalent amount of type 2 diabetic islets did not fully correct hyperglycemia when transplanted into diabetic mice in comparison with normal islets [98]. In another study, when insulin secretion was
Figure 24.7 Preparation of isolated human islets: (a) the pancreas is distended by the injection of a solution containing the digesting enzymes; (b) this leads to the separation of islets (arrows; the red color is due to dithizone staining) from acinar fragments; (c) islet enrichment is achieved by gradient purification, which can be done manually (as in the panel) or automatically; (d) the final islet preparation (dithizone stained) can be used for *in vitro* studies. Source: Images from PM’s laboratory. (For a color version of this figure, please see color plate section.)

Figure 24.8 Perifusion studies with isolated islets showing marked reduction of glucose-stimulated insulin release from type 2 diabetic islets (T2D) as compared with nondiabetic islets (ND). Source: Adapted from Deng and Vatamaniuk 2004 [34]. Reproduced with permission of American Diabetes Association.
measured in response to glucose, glibenclamide, and arginine in isolated nondiabetic and type 2 diabetic islets, again no significant difference in insulin release in response to 3.3 mmol L\(^{-1}\) glucose was observed [36]. However, when challenged with 16.7 mmol L\(^{-1}\) glucose, diabetic islets secreted significantly less insulin than nondiabetic islets. Insulin secretion during glibenclamide and arginine stimulation was also lower from diabetic islets than from control islets; of note, in this series type 2 diabetic islets released more insulin in response to non-glucose stimuli than in response to glucose. Interestingly, similar changes in insulin secretion can be induced in nondiabetic human islets by prolonged exposure to increased levels of free fatty acids (in particular, palmitate) or glucose [31,99–102], mimicking findings reported in \textit{in vivo} studies [103,104]. All these data clearly indicate that type 2 diabetic β-cells have defective insulin secretory behavior, which contributes to the development of hyperglycemia at least in part independently of β-cell mass.

**Disruption of β-cell dialogue in the islet**

It has long been recognized that β-cells dialogue with each other and with the other islet endocrine cells, mainly through autocrine and paracrine communication [105–111]. This is made possible by the peculiar organization of endocrine cells in human islets [3,112]—allowing more than 70% of β-cells to be in contact with non-β-cells [112]—and the abundance of gap junctions (containing connexin proteins) and other cell–cell adhesion complexes [109,113–115]. In addition, endocrine cells could influence one another by the action of the respective hormones via intra-islet blood flow [116,117]. Finally, nonhormonal endocrine cell products (such as ATP and zinc) and neurotransmitters (released either by islet cells or by sympathetic and parasympathetic nerve fibers) influence β-cell function [109–111]. Disruption of these multiple levels of β-cell dialogue in human T2DM has been hypothesized. A well-known feature of T2DM is hyperglucagonemia [118–121], but studies on the relative role of α-cell number and function have given conflicting results. Deng et al. [34] found that the proportion of glucagon-containing cells was significantly increased (52 vs. 35%) in their 14 type 2 diabetic as compared to 14 nondiabetic organ donor pancreases, whereas \textit{ex vivo} glucagon release in isolated islet perfusion studies did not differ. However, in a larger series of 50 type 2 diabetic and 52 nondiabetic autotopic samples, it was shown that α-cells mass was essentially the same in type 2 diabetic (366 mg) and nondiabetic subjects (342 mg), and was not influenced by sex, BMI, or type of diabetes treatment [122]. Likewise, no difference in the number of α cells between type 2 diabetic and nondiabetic cases was shown by electron microscopy evaluation of pancreatic specimens from organ donors [15]. Nevertheless, due to the concomitant reduction of β-cells, in these studies the ratio of α cells to β-cells was significantly higher in diabetic islets [34,122]. This suggests that, on the one hand the reduction of intra-islet insulin in T2DM may contribute to a disproportionately elevated glucagon secretion [121], which could be exacerbated by defective glucose regulation of glucagon secretion (due to impaired metabolic control of the α cell ATP-regulated K\(^{+}\) (KATP) channels [123]); on the other hand, exposure to increased intra-islet glucagon concentrations could have adverse effects on the β-cell [124]. Other features of α cells could influence the β-cell in human T2DM. In fact, subpopulations of α cells have been shown to be able to process proglucagon by proconvertase 1/3 to produce glucagon-like peptide-1 (GLP-1) [125,126]. In addition, \textit{ex vivo} active GLP-1 release is higher in islets from type 2 diabetic individuals [126], and recent data show the presence of dipeptidylpeptidase-4 (DPP-4, the GLP-1 inactivating enzyme) in human α cells (Figure 24.9), with reduced DPP-4 activity in type 2 diabetic islets [127]. These findings are in line with the concept that, under given circumstances human α cells could modulate their phenotype to produce GLP-1 in an attempt to protect the endangered β-cells [125–129]; this phenomenon could include α-cell production of glucose-dependent insulinotropic polypeptide, GIP [130]. This novel information is relevant to the controversial issue of the role of incretin-based therapy in T2DM [131–135].

Recent observations have pointed to yet another abnormality of T2DM islets, namely, altered endocrine cell arrangement. Especially in larger islets, architecture is sparse, and spatial cell organization appears more dispersed [10,15]. Given the importance of the above described cell-to-cell communication for normal islet cell function [105–115], this morphologic feature may have physiologic impact. Interestingly, work has shown that connexin-36 (Cx36, a key component of gap junctions) correlates with insulin gene expression in both nondiabetic and type 2 diabetic islets [136]. In addition, exposure of human islets to lipotoxic conditions reduces Cx36 expression, disrupts islet cell architecture, and decreases glucose-stimulated insulin secretion [137].

**Incretin defect**

Oral glucose elicits a greater insulin response than does intravenous glucose. As first demonstrated by Nauck et al. [138], when tested at matched plasma glucose concentrations (isoglycemic protocol) the oral route of glucose administration induces a robust increment in insulin secretion as compared with the intravenous route. In nondiabetic subjects, such potentiation—named incretin effect—averages 50–70% with a standard (75 g) oral glucose load, and increases with the size of the glucose load [134,139]. As depicted in Figure 24.10, the incretin effect is characteristically lost in patients with T2DM [138,140–146]. Recent work using mathematical modeling of β-cell function has made it possible to separate glucose-mediated from incretin-mediated potentiation by combined analysis of oral and intravenous isoglycemic protocols [135]. The results show that, while glucose potentiation is actually augmented...
Figure 24.9 DPP-4 is present in human islet cells, co-localizing with glucagon (a), but not with insulin (b) or somatostatin (c). Source: Adapted from Omar et al. 2014 [127]. Reproduced with permission of Springer Science and Business Media. (For a color version of this figure, please see color plate section.)

Figure 24.10 The incretin effect, measured as the difference (dark shaded areas) between the insulin secretory response to oral and intravenous (i.v.) glucose under isoglycemic conditions (top panels). Note the absence of a significant incretin effect in patients with type 2 diabetes (T2D). Source: Redrawn from data in Muscelli et al. 2008 [146]. (For a color version of this figure, please see color plate section.)
in type 2 diabetic patients compared to nondiabetic subjects (due to the hyperglycemia), incretin potentiation is severely compromised (Figure 24.11). Relevant in this regard is collated evidence documenting that the secretory response of GLP-1—the main incretin hormone—is quite variable in T2D, ranging from reduced to normal, and is poorly correlated with the extent of the incretin defect [133]. Thus, other factors, hormonal or neural, must contribute to the incretin defect of T2DM. Among the hormonal factors, glucose-dependent insulinoactive polypeptide (GIP) causes little potentiation of insulin release in T2DM despite raised GIP responses to oral stimuli [140]. Importantly, the incretin defect of T2DM, which can also be detected in youth [141] and in MODY-3 (HNF-1A) diabetes [142], is not reversed by treatment with a DPP-4 [143] or metformin [144]. Importantly, the incretin defect of T2DM, which can also be detected in youth [141] and in MODY-3 (HNF-1A) diabetes [142], is not reversed by treatment with a DPP-4 [143] or metformin [144]. These observations have led investigators to postulate that loss of incretin effect is an inherent, rather than secondary, characteristic of T2DM [145].

An important observation is that obesity per se is associated with an impaired incretin effect, that is, independently of glucose tolerance [146]. Because the insulin resistance of obesity involves adipose tissue as well as skeletal muscle, elevated circulating free fatty acids may be linked with the incretin defect, possibly via the reduction of Cx36 [137].

**Additional aspects of β-cell dysfunction**

Detailed analysis of β-cell function in vivo has shown that the ability of β-cells to respond to the rate of increase in plasma glucose concentration (rate sensitivity)—a response modality first identified in the isolated rat pancreas [147]—is markedly impaired in patients with T2DM [80], and is improved by bariatric surgery [87]. Furthermore, the autocrine effect of insulin to stimulate its own release has been shown to be significantly reduced in insulin-resistant states [148]. With regard to the relationship between insulin resistance and β-cell function, separate measurement of insulin sensitivity (by the euglycemic hyperinsulinemic clamp technique) and β-cell function (as resolved by mathematical modeling of C-peptide responses to oral glucose or mixed meals) has demonstrated that these two key determinants of glucose levels are numerically independent of each other across a range of glucose tolerance [25,30] but are nevertheless co-present in diabetes and prediabetes, in adults [80] as well as children [141]. Recent studies in patients undergoing pancreatectoduodenectomy [149] have shed some light on this issue. In fact, insulin-resistant subjects showed an increased islet size and an elevated number of β-cells as well as α-cells, resulting in an reduced β-cell-to-α-cell area, as compared to insulin-sensitive subjects. Following surgical excision of comparable portions of pancreas, the insulin response to intravenous glucose and to a mixed meal was unchanged in the insulin-sensitive individuals whereas it was markedly impaired in the insulin-resistant group.

Finally, impaired β-cell glucose sensitivity is a powerful predictor of progression to dysglycemia and T2DM in subjects with normal glucose tolerance independently of insulin resistance and on top of the classical phenotypic indicators (age, adiposity, familial diabetes) [150]. In multivariate predictive

---

**Figure 24.11** Time-course of glucose-induced and incretin-induced potentiation of insulin secretion in nondiabetic subjects (NGT) and in patients with type 2 diabetes. Source: Redrawn from Tura et al. 2014 [135]. Reproduced with permission of Springer Science and Business Media. (For a color version of this figure, please see color plate section.)
models, absolute insulin secretion is a positive antecedent of deteriorating glucose tolerance, thereby emphasizing the contrasting value of absolute insulin release viz the dynamics of insulin response.

**Loss of β-cell functional identity**

The dominance of β-cell functional impairment in T2DM implies that β-cells have lost, at least in part, their normal insulin secretory phenotype. The associated molecular features have been discussed in a number of insightful reviews and research articles, and the role of genetic, epigenetic, transcriptomic and proteomic changes has been described extensively [151–166]. At the cellular level, β-cell insulin degranulation and the recently hypothesized β-cell dedifferentiation phenomenon could play key roles. Insulin granules can be easily identified by electron microscopy on the basis of their typical morphology, characterized by a dense core and a more or less clearly visible halo [15,36,167,168]. In addition, secretory granules can be subdivided into mature and immature, based on distinct ultrastructural, biochemical, and functional characteristics [167–171], with a relative ratio of 6 to 10 in normal human β-cells [18,172]. Notably, this ratio may be under genetic control, as documented by the finding that β-cells from nondiabetic carriers of the Gly(972)→Arg IRS-1 polymorphism contain a several-fold greater number of immature secretory granules and a lower number of mature granules compared to control β-cells [173]. Also, depending on the technique used, it has been estimated that the number of insulin granules in rodent β-cells ranges from 5000 to more than 10,000 per cell [168,174], of which around 10% are in close proximity to the plasma membrane (“docked” granules) [175]. A few studies have suggested that docked granules could be mainly involved in the first phase of insulin release, but this issue remains controversial [161–171,176–178].

The amount of insulin granules in β-cells from human subjects with T2DM has been investigated by electron microscopy, showing an average 30–40% reduction versus control samples, due to a lower number of mature granules [18,172]. Interestingly, the amount of docked granules has been reported to be similar in type 2 diabetic and nondiabetic β-cells, but with a much higher proportion of immature granules (around fourfold compared to the control cases) [172]. In addition to these changes, decreased gene and/or protein expression of molecules involved in vesicle exocytosis (such as syntaxin-1A, SNAP-25, VAMP-2, Munc-18, Munc 13-1, synaptotagmin V and synaptophysin) has been found in islets from type 2 diabetic patients [179]. Clearly, these changes can conceivably contribute to impair insulin turnover in β-cells from diabetic subjects [180]. At the same time, the insulin granule density of human type 2 diabetic β-cells can be restored (Figure 24.12). In fact, when islets prepared from type 2 diabetic donors were studied after a 24-hour incubation in culture medium containing 5.5 mM glucose and therapeutic concentration of metformin, the volume of mature insulin granules increased from 1.9 ± 0.5 to 3.4 ± 0.3%, which was similar to the value in nondiabetic β-cells (3.2 ± 0.7%) [18]. Although further work is needed, these latter results are in agreement with observations showing replenishment of β-cell granules after reduction of blood glucose in the diabetic *Psammomys obesus* [181], and with *ex vivo* and *in vivo* experiments demonstrating recovery of insulin secretory function of human type 2 diabetic β-cells [26,88,165,182–184].

Degranulation can also lead to underestimation of β-cell number. In a recent article [15], pancreas samples from nondiabetic and matched type 2 diabetic subjects were studied in parallel by immunocytochemistry and electron microscopy. In addition, morphologic, ultrastructural and glucose-stimulated insulin secretion experiments were performed with nondiabetic islets after 24 hours in culture in the presence of 22.2 mM glucose [15]. By immunocytochemistry, the fractional islet

![Figure 24.12](image-url)
Insulin-positive area was 30–40% lower in type 2 diabetic islets, as expected. However, electron microscopy showed that the amount of β-cells in the diabetic islets was only marginally decreased, but marked β-cell degranulation was evident. These findings were reproduced after exposing nondiabetic islets to high glucose. It is therefore possible that a proportion of β-cells in T2DM islets may not be detectable by standard immunohistochemistry staining due to insulin degranulation, potentially leading to an overestimation of β-cell loss. At the same time, the study confirmed reduced glucose-stimulated insulin secretion from both type 2 diabetic islets and nondiabetic islets exposed to high glucose, indicating that major β-cell secretory impairment may indeed occur without any actual loss of β-cells [15].

Interestingly, some islet cells expressing insulin may also show glucagon immunoreactivity [12,149]. In a study, it has been reported that the percentage of insulin-positive cells that were also glucagon immunoreactive is greater in human T2DM (3.2 ± 1.4%) than in nondiabetic samples (0.4 ± 0.1%), and that this proportion may be much higher (16.8 ± 5.0%) depending on the type of antidiabetic therapy [12]. The presence of cells expressing both insulin and glucagon has been also confirmed by gene expression studies on “β-cells” obtained by the laser capture microdissection technique [161], and could be interpreted as a potential β-cell transdifferentiation to α cells in response to insulin resistance [149]. The secretory phenotype of these dual hormone-positive cells is currently unknown. However, cells containing both insulin and glucagon may also be considered as β-cells at some stage of differentiation. In rodent experiments deleting the key transcription factor FOXO1 in combination with permanent genetic labeling of β-cells, it has been shown that the apparent loss of β-cells was not due to their death, but to the missed expression of fundamental β-cell genes, including insulin, glucose transporter 2, and glucokinase [185]. Furthermore, a proportion of FOXO1-deficient β-cells became able to express glucagon [185], all this suggesting that under given conditions insulin degranulation could be a step towards more profound β-cell changes, such as dedifferentiation and transdifferentiation [186–188] (see Figure 24.13). Although some evidence is available to suggest that β-cells dedifferentiation may also occur in human T2DM [188,189], clearly more work is needed to evaluate if and to what extent dedifferentiation has a role in the loss of β-cell functional identity in the disease. If this was the case, new possibilities could be envisaged to protect and/or rescue the β-cell. For instance, resveratrol administration prevented β-cell dedifferentiation in the Rhesus monkey on a diabetogenic diet [190], and insulin therapy with restoration of near-normal glycaemia was associated with β-cell redifferentiation in a mouse model of insulin-secretory deficiency by β-cell inexcitability [191].

Conclusions

β-cell failure is central to the onset and progression of human T2DM. Both decreased β-cell mass and impaired insulin secretory function have been documented ex vivo and in vivo, due to inherited and acquired factors, leading to varying combinations of molecular and cellular alterations typical of the disease. Growing evidence indicates that functional defects are likely to play the dominant role in most cases of diabetes, with a proportion of β-cells escaping death at the expense of the acquisition of a non-functional phenotype. Protection and rescue of the β-cell insulin secretory function should represent the primary objective of any intervention aimed at the prevention and treatment of T2DM.

References


Marchetti P, Del Guerra S, Marselli L, et al.: Pancreatic islets from type 2 diabetic patients have functional defects and increased apoptosis that are ameliorated by metformin. *Journal of Clinical Endocrinology & Metabolism* 2004;89:5535–5541.


Chapter 24


CHAPTER 25

Pathogenesis of type 2 diabetes mellitus

Ralph A. DeFronzo
Diabetes Division, University of Texas Health Science Center, San Antonio, TX, USA

Key points

- Type 2 diabetes is characterized by multiple pathophysiologic disturbances including:
  - muscle insulin resistance → decreased/glucose uptake
  - hepatic insulin resistance → increased/glucose production
  - adipocyte insulin resistance → increase plasma FFA and increased insulin resistance provoking adipokines
  - progressive β-cell failure
  - hyperglucagonemia and increased hepatic sensitivity to glucagon
  - impaired incretin (GLP-1 and GIP) effect
  - increased renal glucose reabsorption
  - brain neurotransmitter dysfunction leading to impaired appetite suppression and weight gain.

- Insulin resistance is the earliest detectable abnormality in the natural history of type 2 diabetes.

- With progressive β-cell failure, in the presence of insulin resistance, individuals progress from normal glucose tolerance to impaired glucose tolerance to overt type 2 diabetes.

Normal glucose homeostasis

Any discussion of the pathogenesis of type 2 diabetes mellitus (T2DM) must start with a review of mechanisms involved in the maintenance of normal glucose homeostasis in the basal or postabsorptive state (10–12-h overnight fast) and following ingestion of a typical mixed meal [1–4]. In the postabsorptive state the great majority of total body glucose disposal takes place in insulin independent tissues. Thus, ~50% of all glucose utilization occurs in the brain, which is insulin independent and becomes saturated at a plasma glucose concentration of about 40 mg dL$^{-1}$ [4,5]. Another 25% of glucose disposal occurs in the splanchnic area (liver plus gastrointestinal tissues) and is also insulin independent. The remaining 25% of glucose utilization in the postabsorptive state takes place in insulin-dependent tissues, primarily muscle, and to a lesser extent adipose tissue. Basal glucose utilization, ~2.0 mg kg$^{-1}$ min$^{-1}$, is precisely matched by the rate of endogenous glucose production (Figure 25.1). Approximately 85% of endogenous glucose production is derived from the liver, and the remaining 15% is produced by the kidney. Glycogenolysis and gluconeogenesis contribute equally to the basal rate of hepatic glucose production.

Following glucose ingestion, the increase in plasma glucose concentration stimulates insulin release, and the combination of hyperinsulinemia and hyperglycemia (i) stimulates glucose uptake by splanchnic (liver and gut) and peripheral (primarily muscle) tissues (Table 25.1), and (ii) suppresses endogenous (primarily hepatic) glucose production [1 –4,6–10]. The majority (~80–85%) of glucose uptake by peripheral tissues occurs in muscle, with a small amount (~4–5%) being metabolized by adipocytes. Although fat tissue is responsible for only a small amount of total body glucose disposal, it plays a very important role in the maintenance of normal glucose homeostasis by regulating the release of free fatty acids (FFA) from stored triglycerides (see later) and through the production of adipocytokines that influence insulin sensitivity in muscle and liver [11–13]. Insulin is a potent antilipolytic hormone and even small increments in the plasma insulin concentration markedly inhibit lipolysis, leading to a decline in the plasma FFA level [12]. The decline in plasma FFA concentration in response to increased plasma levels of insulin and glucose play an important role in the maintenance of normal glucose homeostasis [11–13].

Glucagon also plays a central role in the regulation of glucose homeostasis [14,15]. Under postabsorptive conditions approximately half of total hepatic glucose output is dependent upon the maintenance of normal basal glucagon levels and inhibition of basal glucagon secretion with somatostatin causes a reduction in hepatic glucose production and plasma glucose concentration. After a glucose-containing meal, glucagon secretion is inhibited by hyperinsulinemia, and the resultant
hypoglucagonemia contributes to the suppression of hepatic glucose production and maintenance of normal postprandial glucose tolerance.

The route of glucose entry into the body also plays an important role in the distribution of administered glucose and overall glucose homeostasis [14,16,17]. Intravenous (i.v.) insulin exerts only a small stimulatory effect on splanchnic (liver plus gut) glucose uptake, while i.v. glucose augments splanchnic glucose uptake in direct proportion to the increase in plasma glucose concentration [6]. In contrast, oral glucose administration markedly enhances splanchnic glucose uptake. The portal signal that stimulates hepatic glucose uptake after glucose ingestion is directly proportional to the negative hepatic artery-portal vein glucose concentration gradient [9]. As this gradient widens, the splanchnic nerves are stimulated and this activates a neural reflex, in which vagal activity is enhanced and sympathetic nerves innervating the liver are inhibited. These neural changes augment liver glucose uptake and stimulate hepatic glycogen synthase, while simultaneously inhibiting glycogen phosphorylase. In contrast to i.v. glucose/insulin administration, where muscle accounts for the majority (≈80–85%) of glucose disposal, the splanchnic tissues are responsible for the removal of ≈30–40% of an ingested glucose load.

Glucose administration via the gastrointestinal tract also has a potentiating effect on insulin secretion. Thus, the plasma insulin response following oral glucose is about twice as great as that following i.v. glucose despite equivalent increases in the plasma glucose concentration. This incretin effect is related to the release of glucagon-like peptide-1 (GLP-1) from the L cells in the large intestine and glucose-dependent insulinotropic polypeptide (previously called gastric inhibitory polypeptide) (GIP) from the K cells in the early part of the small intestine [18–20]. GLP-1 also inhibits glucagon secretion and the decline in plasma glucagon concentration contributes to suppression of hepatic glucose production following meal ingestion. GLP-1 and GIP are released in response to nutrient absorption by the L and K cells [18–20], and in response to a meal increased circulating levels of GLP-1 and GIP can be detected within minutes, before nutrients can reach the K cells in the duodenum and long before they reach the L cells in the distal small intestine and early large intestine. This early release of GLP-1 and GIP is mediated via neural impulses that are carried to centers in the hypothalamus and back to the intestinal cells via the vagus nerve [21]. GLP-1 and GIP have their own receptors on the β cell and augment insulin secretion by activation of adenyly cyclase [18–20]. Importantly, the stimulation of insulin secretion by GLP1 and GIP is glucose-dependent, that is, insulin release is augmented in the presence of hyperglycemia and the stimulatory effect of both GLP-1 and GIP wanes as the blood glucose concentration returns to normoglycemic levels.

### Natural history of type 2 diabetes

The natural history of T2DM has been well described in multiple populations [1,3,22–32] and is reviewed in references [1] and [3]. Individuals destined to develop T2DM inherit a set of genes from their parents that make their tissues resistant to insulin [1–3,33–37] and the insulin resistance is aggravated by weight gain and physical inactivity. Hepatic insulin resistance is manifested by an overproduction of glucose during the basal state despite the presence of fasting hyperinsulinemia [38] and an impaired suppression of hepatic glucose production (HGP) in response to insulin [39], as occurs following a meal [40]. Muscle insulin resistance [6,36,39,41] is manifest by impaired glucose uptake following ingestion of a carbohydrate meal and results in postprandial hyperglycemia [40]. The origin of the insulin resistance can be traced to the genetic background [2,34,42]. However, the epidemic of diabetes that has enveloped Westernized countries primarily results from the epidemic of obesity and physical inactivity [43]. Both obesity [44] and decreased physical activity [45] are insulin-resistant states and, when added to the genetic burden of the insulin resistance, place a major stress on the pancreatic β cells to augment their secretion of insulin to offset the defect in insulin action [1–3]. Initially the β cells augment their secretion of insulin to offset the insulin resistance and glucose tolerance remains normal [46]. However, with time the β cells begin to fail and initially the postprandial plasma glucose levels and subsequently the
fasting plasma glucose concentration rise, leading to the onset of overt diabetes [1–3,47,48]. Collectively, the insulin resistance in muscle and liver and β-cell failure have been referred to as the Triumvirate [1] (Figure 25.1). The resultant hyperglycemia and poor metabolic control may cause a further decline in insulin sensitivity, but it is the progressive β-cell failure that determines the rate of disease progression.

Although the relative contributions of insulin resistance and β-cell failure to the development of T2DM may vary amongst different ethnic groups [49], progressive β-cell failure superimposed upon a background of insulin resistance represent the core pathophysiologic defects responsible for the development of overt diabetes [1–3,50].

The natural history of T2DM described above [1–3] is depicted by a prospective 6-year study carried out by Felber and colleagues [26] (Figure 25.2). Subjects had a euglycemic clamp to measure tissue sensitivity to insulin and an oral glucose tolerance test (OGTT) to provide a measure of glucose clamp to measure insulin sensitivity. With time, the high rate of insulin secretion cannot be maintained, the β cell starts on the downward slope of Starling’s curve and fasting hyperglycemia and glucose intolerance ensue.

**β-Cell function**

Although the plasma insulin response to the development of insulin resistance typically is increased during the natural history of T2DM (Figure 25.2), this does not mean that the β cell is functioning normally. To the contrary, studies have demonstrated that the onset of β-cell failure occurs much earlier and is more severe than previously appreciated. In the San Antonio Metabolism (SAM) study and the Veterans Administration Genetic Epidemiology Study (VAGES), a large number of subjects with NGT (n = 318), IGT (n = 259), and T2DM (n = 201) were studied [55–58]. All subjects had an OGTT with plasma glucose and insulin concentrations measured every 15 min to evaluate overall glucose tolerance and β-cell function and a euglycemic insulin clamp to measure insulin sensitivity. Simply measuring the plasma insulin response to a glucose challenge does not provide a valid index of β-cell function [59]. The β cell responds to an increment in glucose (ΔG) with an increment in insulin (ΔI) [59]. Thus, a better measure of β-cell function is ΔI/ΔG. However, the β cell also recognizes the severity of insulin resistance and adjusts its secretion of insulin to offset the defect in insulin action [46,59–61]. Thus, the gold standard for measuring β-cell function is the insulin secretion/insulin resistance (β-Cell function)

**β-Cell function**

Although the plasma insulin response to the development of insulin resistance typically is increased during the natural history of T2DM (Figure 25.2), this does not mean that the β cell is functioning normally. To the contrary, studies have demonstrated that the onset of β-cell failure occurs much earlier and is more severe than previously appreciated. In the San Antonio Metabolism (SAM) study and the Veterans Administration Genetic Epidemiology Study (VAGES), a large number of subjects with NGT (n = 318), IGT (n = 259), and T2DM (n = 201) were studied [55–58]. All subjects had an OGTT with plasma glucose and insulin concentrations measured every 15 min to evaluate overall glucose tolerance and β-cell function and a euglycemic insulin clamp to measure insulin sensitivity. Simply measuring the plasma insulin response to a glucose challenge does not provide a valid index of β-cell function [59]. The β cell responds to an increment in glucose (ΔG) with an increment in insulin (ΔI) [59]. Thus, a better measure of β-cell function is ΔI/ΔG. However, the β cell also recognizes the severity of insulin resistance and adjusts its secretion of insulin to offset the defect in insulin action [46,59–61]. Thus, the gold standard for measuring β-cell function is the insulin secretion/insulin resistance (β-Cell function)

**β-Cell function**

Although the plasma insulin response to the development of insulin resistance typically is increased during the natural history of T2DM (Figure 25.2), this does not mean that the β cell is functioning normally. To the contrary, studies have demonstrated that the onset of β-cell failure occurs much earlier and is more severe than previously appreciated. In the San Antonio Metabolism (SAM) study and the Veterans Administration Genetic Epidemiology Study (VAGES), a large number of subjects with NGT (n = 318), IGT (n = 259), and T2DM (n = 201) were studied [55–58]. All subjects had an OGTT with plasma glucose and insulin concentrations measured every 15 min to evaluate overall glucose tolerance and β-cell function and a euglycemic insulin clamp to measure insulin sensitivity. Simply measuring the plasma insulin response to a glucose challenge does not provide a valid index of β-cell function [59]. The β cell responds to an increment in glucose (ΔG) with an increment in insulin (ΔI) [59]. Thus, a better measure of β-cell function is ΔI/ΔG. However, the β cell also recognizes the severity of insulin resistance and adjusts its secretion of insulin to offset the defect in insulin action [46,59–61]. Thus, the gold standard for measuring β-cell function is the insulin secretion/insulin resistance (β-Cell function)

**β-Cell function**

Although the plasma insulin response to the development of insulin resistance typically is increased during the natural history of T2DM (Figure 25.2), this does not mean that the β cell is functioning normally. To the contrary, studies have demonstrated that the onset of β-cell failure occurs much earlier and is more severe than previously appreciated. In the San Antonio Metabolism (SAM) study and the Veterans Administration Genetic Epidemiology Study (VAGES), a large number of subjects with NGT (n = 318), IGT (n = 259), and T2DM (n = 201) were studied [55–58]. All subjects had an OGTT with plasma glucose and insulin concentrations measured every 15 min to evaluate overall glucose tolerance and β-cell function and a euglycemic insulin clamp to measure insulin sensitivity. Simply measuring the plasma insulin response to a glucose challenge does not provide a valid index of β-cell function [59]. The β cell responds to an increment in glucose (ΔG) with an increment in insulin (ΔI) [59]. Thus, a better measure of β-cell function is ΔI/ΔG. However, the β cell also recognizes the severity of insulin resistance and adjusts its secretion of insulin to offset the defect in insulin action [46,59–61]. Thus, the gold standard for measuring β-cell function is the insulin secretion/insulin resistance (β-Cell function)

**β-Cell function**

Although the plasma insulin response to the development of insulin resistance typically is increased during the natural history of T2DM (Figure 25.2), this does not mean that the β cell is functioning normally. To the contrary, studies have demonstrated that the onset of β-cell failure occurs much earlier and is more severe than previously appreciated. In the San Antonio Metabolism (SAM) study and the Veterans Administration Genetic Epidemiology Study (VAGES), a large number of subjects with NGT (n = 318), IGT (n = 259), and T2DM (n = 201) were studied [55–58]. All subjects had an OGTT with plasma glucose and insulin concentrations measured every 15 min to evaluate overall glucose tolerance and β-cell function and a euglycemic insulin clamp to measure insulin sensitivity. Simply measuring the plasma insulin response to a glucose challenge does not provide a valid index of β-cell function [59]. The β cell responds to an increment in glucose (ΔG) with an increment in insulin (ΔI) [59]. Thus, a better measure of β-cell function is ΔI/ΔG. However, the β cell also recognizes the severity of insulin resistance and adjusts its secretion of insulin to offset the defect in insulin action [46,59–61]. Thus, the gold standard for measuring β-cell function is the insulin secretion/insulin resistance (β-Cell function)

**β-Cell function**

Although the plasma insulin response to the development of insulin resistance typically is increased during the natural history of T2DM (Figure 25.2), this does not mean that the β cell is functioning normally. To the contrary, studies have demonstrated that the onset of β-cell failure occurs much earlier and is more severe than previously appreciated. In the San Antonio Metabolism (SAM) study and the Veterans Administration Genetic Epidemiology Study (VAGES), a large number of subjects with NGT (n = 318), IGT (n = 259), and T2DM (n = 201) were studied [55–58]. All subjects had an OGTT with plasma glucose and insulin concentrations measured every 15 min to evaluate overall glucose tolerance and β-cell function and a euglycemic insulin clamp to measure insulin sensitivity. Simply measuring the plasma insulin response to a glucose challenge does not provide a valid index of β-cell function [59]. The β cell responds to an increment in glucose (ΔG) with an increment in insulin (ΔI) [59]. Thus, a better measure of β-cell function is ΔI/ΔG. However, the β cell also recognizes the severity of insulin resistance and adjusts its secretion of insulin to offset the defect in insulin action [46,59–61]. Thus, the gold standard for measuring β-cell function is the insulin secretion/insulin resistance (β-Cell function)

**β-Cell function**

Although the plasma insulin response to the development of insulin resistance typically is increased during the natural history of T2DM (Figure 25.2), this does not mean that the β cell is functioning normally. To the contrary, studies have demonstrated that the onset of β-cell failure occurs much earlier and is more severe than previously appreciated. In the San Antonio Metabolism (SAM) study and the Veterans Administration Genetic Epidemiology Study (VAGES), a large number of subjects with NGT (n = 318), IGT (n = 259), and T2DM (n = 201) were studied [55–58]. All subjects had an OGTT with plasma glucose and insulin concentrations measured every 15 min to evaluate overall glucose tolerance and β-cell function and a euglycemic insulin clamp to measure insulin sensitivity. Simply measuring the plasma insulin response to a glucose challenge does not provide a valid index of β-cell function [59]. The β cell responds to an increment in glucose (ΔG) with an increment in insulin (ΔI) [59]. Thus, a better measure of β-cell function is ΔI/ΔG. However, the β cell also recognizes the severity of insulin resistance and adjusts its secretion of insulin to offset the defect in insulin action [46,59–61]. Thus, the gold standard for measuring β-cell function is the insulin secretion/insulin resistance (β-Cell function)

**β-Cell function**

Although the plasma insulin response to the development of insulin resistance typically is increased during the natural history of T2DM (Figure 25.2), this does not mean that the β cell is functioning normally. To the contrary, studies have demonstrated that the onset of β-cell failure occurs much earlier and is more severe than previously appreciated. In the San Antonio Metabolism (SAM) study and the Veterans Administration Genetic Epidemiology Study (VAGES), a large number of subjects with NGT (n = 318), IGT (n = 259), and T2DM (n = 201) were studied [55–58]. All subjects had an OGTT with plasma glucose and insulin concentrations measured every 15 min to evaluate overall glucose tolerance and β-cell function and a euglycemic insulin clamp to measure insulin sensitivity. Simply measuring the plasma insulin response to a glucose challenge does not provide a valid index of β-cell function [59]. The β cell responds to an increment in glucose (ΔG) with an increment in insulin (ΔI) [59]. Thus, a better measure of β-cell function is ΔI/ΔG. However, the β cell also recognizes the severity of insulin resistance and adjusts its secretion of insulin to offset the defect in insulin action [46,59–61]. Thus, the gold standard for measuring β-cell function is the insulin secretion/insulin resistance (β-Cell function)
Not be equated with T2DM, according to current diagnostic criteria [63]. Multiple variables are strongly and linearly related \((r = 0.91, p < 0.00001)\). There are no cut points that distinguish NGT from IGT or IGT from T2DM. Rather, glucose intolerance is a continuum, and subjects simply move up and down this curve as a function of the insulin secretion/insulin resistance index. Therefore, the current diagnostic criteria [63] for IGT and T2DM are quite arbitrary and glucose tolerance should be viewed as a continuum of risk. The higher the 2-h plasma glucose concentration, even within the range of IGT, the greater is the risk for microvascular complications.

The previous discussion has focused on \(\beta\)-cell function and clearly demonstrates that \(\beta\)-cell health is severely impaired well before the onset of T2DM and even before the development of IGT. Even more ominous are studies demonstrating a significant reduction in \(\beta\)-cell mass in prediabetic (IFG/IGT) individuals [64–66]. In a postmortem analysis, these investigators demonstrated that as individuals progress from NGT to impaired fasting glucose (IFG), there was a 50% decline in \(\beta\)-cell volume, suggesting a significant loss of \(\beta\)-cell mass long before the onset of T2DM. With the progression to overt diabetes, there was a further loss of \(\beta\)-cell volume. Although \(\beta\)-cell volume should not be equated with \(\beta\)-cell mass, these results indicate that significant loss of \(\beta\)-cell mass occurs long before the onset of T2DM, according to current diagnostic criteria [63]. Multiple other studies have also demonstrated a significant loss of \(\beta\)-cell mass before the onset of diabetes with a further decrease in \(\beta\)-cell mass with progression to overt diabetes [64–68].

In summary, at the upper tertile of IGT [55–58,62], individuals have lost over 80% of their \(\beta\)-cell function, while the results of autopsy studies [64–66] indicate that subjects with "prediabetes" have experienced a significant loss of \(\beta\)-cell mass. In humans this presents a major problem since no therapeutic intervention has been shown to increase \(\beta\)-cell number.

**Type 2 diabetes with hypoinsulinemia**

A large body of clinical and experimental evidence documents that hyperinsulinemia and insulin resistance precede the onset of T2DM. Nonetheless, a number of studies have shown that absolute insulin deficiency, with or without impaired tissue insulin sensitivity, can lead to the development of T2DM. This scenario is best exemplified by patients with maturity onset diabetes of youth (MODY) [69–71]. This familial subtype of T2DM is characterized by early age of onset, autosomal dominant inheritance with high penetrance, mild-to-moderate fasting hyperglycemia, and impaired insulin secretion.

MODY originally was described by Fajans and subsequently it was demonstrated that MODY-1 resulted from a nonsense mutation in exon 7 of the hepatic nuclear factor (HNF4\(\alpha\)) gene [72]. It later was demonstrated that MODY in French families resulted from mutations in the glucokinase gene on chromosome 7p (MODY-2) [73]. More than eight specific mutations in different genes have been implicated in the MODY profile including glucokinase and seven transcription factors [69–74]: MODY-1 = HNF4\(\alpha\); MODY-2 = glucokinase; MODY-3 = HNF1\(\alpha\); MODY-4 = insulin promoter factor 1; MODY-5 = HNF1\(\beta\); MODY-6 = neurogenic differentiation 1/\(\beta\)-cell E-box transactivator 2; MODY-7 = KLF11 or Kruppel-like factor 11 that regulates Pdx1 transcription in \(\beta\) cells; MODY-8 = carboxyl-ester lipase gene. HNF1\(\alpha\), HNF1\(\beta\), and HNF4\(\alpha\) constitute part of a network of transcription factors that function collectively during embryonic development and during adulthood to regulate the expression of the insulin gene. The hallmark defect in MODY individuals is impaired insulin secretion in response to glucose and other secretagogues. However, peripheral tissue resistance to insulin and abnormalities in hepatic glucose metabolism have also been shown to play some role in the development of impaired glucose homeostasis [75]. Although glucokinase mutations are characteristic of MODY-2, genetic studies in typical older-onset type 2 diabetic individuals have shown that glucokinase mutations account for less than 1% of the common form of T2DM [76].

Cerasi, Luft, Hales, and coworkers [77–79] have championed the view that insulin deficiency represents the primary defect responsible for glucose intolerance in typical type 2 diabetic individuals who do not have glucokinase or other MODY mutations. According to these investigators, impaired early
insulin secretion leads to an excessive rise in plasma glucose concentration and the resultant hyperglycemia is responsible for late hyperinsulinemia. Hales and colleagues [78] have demonstrated that many lean Caucasians with mild fasting hyperglycemia (<140 mg dL\(^{-1}\), 7.8 mmol L\(^{-1}\)) are characterized by insulin deficiency at all time points during an OGTT. An impaired early insulin response has also been a characteristic finding in Japanese Americans who progress to T2DM [79]. Unfortunately, none of these studies provided information about insulin sensitivity. In Caucasians several groups [80,81] have demonstrated normal insulin sensitivity in a minority of type 2 diabetic individuals and it has been suggested that up to 50% of African American type 2 diabetic patients who reside in New York City are characterized by severely impaired insulin secretion and normal insulin sensitivity [82]. A similar defect in insulin secretion has been described in black African type 2 diabetic individuals living in Cameroon [83].

In summary, it is clear that impaired insulin secretion — in the absence of insulin resistance — can lead to the development of full-blown T2DM. However, it remains to be clarified how frequently a pure \(\beta\)-cell defect results in typical T2DM in the general population.

**First-phase insulin secretion**

In response to i.v. glucose, insulin is secreted in a biphasic pattern with an early burst of insulin release within the first 10 min followed by a progressively increasing phase of insulin secretion that persists as long as the hyperglycemic stimulus is present [84]. This biphasic insulin response is not observed after oral glucose, because of the more gradual rise in plasma glucose concentration. Loss of first-phase insulin secretion is a characteristic and early abnormality in patients destined to develop T2DM [1–3]. In most type 2 diabetic subjects a reduction in the early phase of insulin secretion during the OGTT (0–30 min) and during the IVGTT (0–10 min) becomes evident when fasting plasma glucose concentration exceeds \(110–120\) mg dL\(^{-1}\) (6.1–6.7 mmol L\(^{-1}\)) [1–3,48,85,86]. During the OGTT, the defect in early insulin secretion is most obvious if the incremental plasma insulin response at 30 min is expressed relative to the incremental plasma glucose response at 30 min \((\Delta I_{30}/\Delta G_{30})\). Although the first-phase insulin secretory response to i.v. glucose characteristically is diminished or lost in T2DM, this defect is not consistently observed until the fasting plasma glucose concentration rises to \(\sim115–120\) mg dL\(^{-1}\) (6.4–6.7 mmol L\(^{-1}\)). The defect in first-phase insulin response can be partially restored with tight metabolic control [87–91], indicating that at least part of the defect is acquired (see subsequent discussion). Loss of the first phase of insulin secretion has important pathogenic consequences, since this early burst of insulin primes insulin target tissues, especially the liver, that are responsible for the maintenance of normal glucose homeostasis [60,61].

**Pathogenesis of \(\beta\)-cell failure (Figure 25.4)**

*Age.* Numerous studies [92,93] have demonstrated a progressive age-related decline in \(\beta\)-cell function. This is consistent with the well-established observation that the incidence of diabetes increases progressively with advancing age. However, it is clear that factors in addition to age must be involved to account for the major impairment in \(\beta\)-cell function in T2DM.

*Genes.* \(\beta\)-Cell failure also clusters in families, and studies in first-degree relatives of T2DM parents and in twins have provided strong evidence for the genetic basis of the \(\beta\)-cell dysfunction [94–96]. A number of genes have been associated with T2DM in multiple ethnic populations. Of these, the most common are transcription factors associated with \(\beta\)-cell function [34,97–102]. In Finnish families with T2DM impaired insulin secretion is an inherited trait with evidence for a susceptibility locus on chromosome 12 [103]. Of these genes, the transcription factor TCF7L2 is best established [97,98]. Studies by Groop and colleagues [104] have shown that the T-allele of single nucleotide polymorphism rs7903146 of the TCF7L2 gene is associated with impaired insulin secretion in *vivo* and reduced responsiveness to glucagon-like peptide 1 (GLP-1). Both the CT and TT genotypes predict T2DM in multiple ethnic groups [105]. In both the Malmö and Botnia studies, presence of either the CT or TT genotype was associated with a significant reduction in the diabetes-free survival time [104].

TCF7L2 encodes for a transcription factor involved in Wnt signaling, which plays a central role in the regulation of \(\beta\)-cell proliferation and insulin secretion [106]. A number of other transcription factors have also been associated with impaired insulin secretion in T2DM including: GCK, a gene responsible for MODY-2; SLC30A8, a zinc transporter involved in maintaining the appropriate amount of zinc in \(\beta\)-cell secretion granules; KCNJ11 and ABCC8 which encode the subunits of the ATP-sensitive potassium channel; and others [102].

At present there are no known therapeutic interventions that can reverse either the age-related decline or genetic-related factors responsible for impaired insulin secretion. However, there
are a number of causes of β-cell failure that can be reversed or ameliorated.

**Insulin resistance.** Insulin resistance places an increased demand on the β cells to hypersecrete insulin [46] and thus contributes to the progressive β-cell failure in T2DM [107]. Therefore, interventions aimed at enhancing insulin sensitivity are of paramount importance. The precise mechanism(s) via which insulin resistance leads to β-cell failure remain(s) unknown. It is commonly stated that the β cell, by being forced to continuously hypersecrete insulin, eventually wears out. Although simplistic in nature, this explanation lacks a mechanistic cause. Nonetheless, β cell “unloading” with thiazolidinediones in IGT subjects has been shown to markedly enhance β-cell function and reduce the conversion to T2DM [107,108]. An alternate hypothesis, for which considerable evidence exists, is that the etiology of the insulin resistance also is responsible for the β-cell failure. Thus, excess deposition of fat (long chain-fatty acyl CoAs, diacylglycerol, and ceramide) in liver and muscle impairs insulin signaling, causing insulin resistance in these organs, that is, lipotoxicity. Similarly, deposition of fat in the β cell and chronically elevated plasma FFA lead to impaired insulin secretion and β-cell failure [109–112]. Hypersecretion of islet amyloid polypeptide (IAPP) has also been implicated as a cause of the β-cell failure [54,113,114]. IAPP is especially toxic to the β cell in the presence of elevated intracellular fat content [114]. Further, as the IAPP ammulates it coalesces and encroaches upon the β-cell, leading to β-cell destruction [54,67,114,115]. Lastly, studies in the β-cell insulin receptor knock out (BIRKO) mouse [116] and in humans with glucotoxicity. Chronically elevated plasma glucose levels impair β-cell function, and this has been referred to as glucotoxicity [126]. Studies by Rossetti et al. [127] have provided definitive proof of this concept. Partially pancreatectomized diabetic rats are characterized by severe defects in both first- and second-phase insulin secretion compared with control rats. Phlorizin, an inhibitor of renal glucose transport, normalizes the plasma glucose profile without change in any other circulating metabolites and restores both the first and second phases of insulin secretion. In vitro studies with isolated human islets have also demonstrated that chronic exposure to elevated plasma glucose levels impairs insulin secretion [128,129]. In rats, elevation of the mean day-long plasma glucose concentration in vivo by as little as 16 mg dL−1 leads to a marked inhibition of glucose-stimulated insulin secretion in the isolated perfused pancreas [130]. In humans correction of hyperglycemia with insulin [87,89–91,131,132] or inhibitors of renal glucose transport (DeFronzo, unpublished results) reverses the glucotoxic effect of chronic hyperglycemia on the β cells [80–84], leading to improved first- and second-phase insulin secretion, as well as reversal of hepatic and muscle insulin resistance.

**Lipotoxicity.** Lipid deposition in the β cell [109,111,112] and chronic elevation of the plasma FFA concentration impair insulin secretion, and this has been referred to as lipotoxicity. A physiologic elevation of the plasma FFA concentration for as little as 48 hours markedly impairs insulin secretion in genetically predisposed individuals [110] (Figure 25.5). In vivo studies in rodents [120,121] and in vitro studies [122,123] also support an important role for lipotoxicity. Incubation of human pancreatic islets for 48 h with FFA (oleate-to-palmitate ratio 2:1) impairs both the acute and late insulin response, inhibits insulin mRNA expression, and reduces islet insulin content [123]. The peroxisome proliferator–activated receptor (PPAR)γ agonist, rosiglitazone, has been shown to prevent all of these deleterious effects of FFA [123,124]. Consistent with these in vitro observations, both rosiglitazone and pioglitazone markedly improve the insulin secretion/insulin resistance index in vivo in type 2 diabetic humans [125]. Weight loss, which mobilizes fat out of the β cell, also reverses lipotoxicity and preserves β-cell function [109].

**Glucotoxicity.** Chronically elevated plasma glucose levels impair β-cell function, and this has been referred to as glucotoxicity [126]. Studies by Rossetti et al. [127] have provided definitive proof of this concept. Partially pancreatectomized diabetic rats are characterized by severe defects in both first- and second-phase insulin secretion compared with control rats. Phlorizin, an inhibitor of renal glucose transport, normalizes the plasma glucose profile without change in any other circulating metabolites and restores both the first and second phases of insulin secretion. In vitro studies with isolated human islets have also demonstrated that chronic exposure to elevated plasma glucose levels impairs insulin secretion [128,129]. In rats, elevation of the mean day-long plasma glucose concentration in vivo by as little as 16 mg dL−1 leads to a marked inhibition of glucose-stimulated insulin secretion in the isolated perfused pancreas [130]. In humans correction of hyperglycemia with insulin [87,89–91,131,132] or inhibitors of renal glucose transport (DeFronzo, unpublished results) reverses the glucotoxic effect of chronic hyperglycemia on the β cells [80–84], leading to improved first- and second-phase insulin secretion, as well as reversal of hepatic and muscle insulin resistance.

**IAPP.** Excessive secretion of IAPP with subsequent amyloid deposition within the pancreas has also been shown to contribute to progressive β-cell failure in T2DM [54,113,114,133]. Convincing evidence for a pathogenic role of IAPP exists in rodents [134,135] and baboons [54,67,115] and the natural history of pancreatic amylin deposition in humans parallels that in rodents and primates [136].

![Figure 25.5](image-url) Effect of physiologic elevation (48 h) in the plasma FFA concentration (brought about by lipid infusion) on plasma C-peptide concentration (left) and insulin secretory response (deconvolution of the plasma C-peptide curve) (right) in offspring of two type 2 diabetic parents. Source: DeFronzo RA. Diabetes 2003;52:2461–2474.
The baboon genome shares more than 98% homology with the human genome [137,138]. Therefore, results in baboons are likely to be pertinent to those in humans. As the relative amyloid area of the pancreatic islets increase from <5.5% to >51%, there is a progressive decline in the log of HOMA-\(\beta\), which was strongly correlated with the increase in fasting plasma glucose concentration [54]. Studies by the investigators [139,140] have provided additional evidence for a \(\beta\)-cell toxic effect for soluble IAPP fibrils.

It follows that interventions that improve insulin sensitivity, that is, TZDs/metformin/weight loss, by leading to a reduction in insulin secretion, and therefore IAPP secretion (insulin and IAPP are co-secreted in a one-to-one molar ratio), would be expected to preserve \(\beta\)-cell function, and rosiglitazone has been shown to protect human islets against IAPP toxicity by a PI-3 kinase-dependent pathway [141].

**Incretins.** In T2DM some investigators have demonstrated a small decline in GLP-1 secretion or a delayed GLP-1 response (reviewed in [142]), while GIP secretion has been reported to be normal or slightly increased [142]. More importantly, there is severe resistance to the stimulatory effect of both GLP-1 and GIP [143–145]. The resistance to GLP-1 can be observed in individuals with IGT and worsens progressively with progression to T2DM [146]. Of note the resistance to GLP-1 can be overcome by infusing GLP-1 [147] or administration of GLP-1 analogues [148–150] to generate pharmacologic levels (70–90 pM) of the incretin. Tight glycemic control for as little as 4 weeks can improve the \(\beta\) cells’ insulin secretory response to both GLP-1 and GIP [151]. Studies in patients with chronic pancreatitis and T2DM also indicate that the reduced incretin defect in T2DM is not a primary effect for the development of impaired insulin secretion [152]. Thus, \(\beta\)-cell resistance to GLP-1 and GIP is another manifestation of glucotoxicity.

**In utero fetal malnutrition**

Low birth weight is associated with the development of IGT and T2DM in a number of populations [153,154]. Developmental studies in animals and humans have demonstrated that poor nutrition and impaired fetal growth (small babies at birth) are associated with impaired insulin secretion and/or reduced \(\beta\)-cell mass. Fetal malnutrition can also lead to the development of insulin resistance later in life [155]. One could hypothesize that an environmental influence, for example, impaired fetal nutrition leading to an acquired defect in insulin secretion or reduced \(\beta\)-cell mass, when superimposed on insulin resistance, could eventuate in T2DM later in life. Thus, during the normal aging process, with the onset of obesity, or with a worsening of the genetic component of the insulin resistance, the \(\beta\) cell would be called upon to augment its secretion of insulin to offset the defect in insulin action. If \(\beta\)-cell mass (or function) is reduced (or impaired) by an environmental insult during fetal life, this would lead to the development of IGT and eventually overt T2DM. Although such a defect would limit the maximum amount of insulin that could be secreted, it would not explain the progressive decline in insulin secretion in response to physiologic stimuli as individuals progress from NGT to IGT to overt T2DM mellitus.

**Summary.** Although insulin resistance in liver and muscle are well established early in the natural history of the disease, overt T2DM does not occur in the absence of progressive \(\beta\)-cell failure.

**Insulin resistance and type 2 diabetes mellitus**

In cross-sectional studies and long-term, prospective longitudinal studies hyperinsulinemia has been shown to precede the onset of T2DM in all ethnic populations with a high incidence of T2DM [1–3,156–164]. Studies utilizing the euglycemic insulin clamp, minimal model, and insulin suppression techniques have provided direct quantitative evidence that the progression from normal to impaired glucose tolerance is associated with the development of severe insulin resistance, whereas plasma insulin concentrations, both in the fasting state and in response to a glucose load (Figures 25.3 and 25.4) are increased when viewed in absolute terms (see earlier discussion about insulin secretion).

Himsworth and Kerr, using a combined oral glucose and i.v. insulin tolerance test, were the first to demonstrate that tissue sensitivity to insulin was diminished in type 2 diabetic patients [165]. In 1975 Reaven and colleagues, using the insulin suppression test, provided further evidence that the ability of insulin to promote tissue glucose uptake in T2DM was severely reduced [166]. A defect in insulin action in T2DM also has been demonstrated with the arterial infusion of insulin into the brachial artery (forearm muscle) and femoral artery (leg muscle), as well as with radioisotope turnover studies, the frequently sampled i.v. glucose tolerance test, and the minimal model technique [1–3,167–169].

DeFronzo et al., using the more physiologic euglycemic insulin clamp technique, have provided the most conclusive documentation that insulin resistance is a characteristic feature of lean, as well as obese, type 2 diabetic individuals [1–3,6,12,170,171]. Because diabetic patients with severe fasting hyperglycemia (>180–200 mg dL\(^{-1}\), 10.0–11.1 mmol L\(^{-1}\)) are insulinoenic (Figure 25.3), and because insulin deficiency is associated with the emergence of a number of intracellular defects in insulin action, these initial studies focused on diabetics with mild to modest elevations in the fasting plasma glucose concentration (mean = 150 ± 8 mg dL\(^{-1}\), 8.3 ± 0.4 mmol L\(^{-1}\)). Insulin-mediated whole-body glucose disposal in these lean diabetics was reduced by ~40–50%, providing conclusive proof of the presence of moderate to severe insulin resistance. Three additional points are noteworthy: (i) lean type 2 diabetics with more severe fasting hyperglycemia (198 ±10 mg dL\(^{-1}\)) have a severity of insulin resistance that is only slightly (10–20%)
The fasting plasma glucose concentration is significantly greater than in diabetics with mild fasting hyperglycemia; (ii) the defect in insulin action is observed at all plasma insulin concentrations, spanning the physiologic and pharmacologic range (Figure 25.6); (iii) diabetic patients with overt fasting hyperglycemia cannot elicit a normal glucose metabolic response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulating plasma insulin response to even maximally stimulat...
insulin level is very steep, with a half-maximal insulin concentration (ED50) of \(\sim 30–40 \mu\text{U/mL}^{-1}\); (ii) in type 2 diabetic subjects the dose-response curve is shifted rightward, indicating resistance to the inhibitory effect of insulin on hepatic glucose production. However, elevation of the plasma insulin concentration to the high physiologic range (\(\sim 100 \mu\text{U/mL}^{-1}\)) can overcome the hepatic insulin resistance and cause a near normal suppression of HGP; (iii) the severity of the hepatic insulin resistance is related to the level of glycemic control. In type 2 diabetic patients with mild fasting hyperglycemia, an increment in plasma insulin concentration of 100 \(\mu\text{U/mL}^{-1}\) causes a complete suppression of HPG. However, in diabetic subjects with more severe fasting hyperglycemia, the ability of the same plasma insulin concentration to suppress HGP is impaired. These observations indicate that there is an acquired component of hepatic insulin resistance, which becomes progressively worse as the diabetic state decompensates over time.

The kidney possesses all of the gluconeogenesis enzymes required to produce glucose and estimates of the renal contribution to total endogenous glucose production have varied from 5% to 20% \([192,193]\). These varying estimates of the contribution of renal gluconeogenesis to total glucose production are, in large part, related to differences in the methodology employed to measure renal glucose production \([194]\). One study suggests that the basal rate of renal gluconeogenesis is increased in type 2 diabetics and contributes to the elevation in fasting plasma glucose concentration \([195]\). However, studies employing the hepatic vein catheter technique have shown that all of the increase in total body endogenous glucose production (measured with \(3^-\text{H-glucose}\)) in type 2 diabetics can be accounted for by increased hepatic glucose output (measured by the hepatic vein catheter technique) \([6]\).

**Muscle.** Muscle is the major site of insulin-mediated glucose disposal in humans \([1–3,6]\). Using the euglycemic insulin clamp technique \([191]\) in combination with tritiated glucose to measure total body glucose disposal \([1–3,6,19,28,32,39,41,60,61,177,181,196–199]\), it has been conclusively demonstrated that lean type 2 diabetic individuals are severely resistant to insulin compared with age-, weight-, and sex-matched control subjects. Employing femoral arterial and venous catheterization in combination with the insulin clamp, muscle insulin resistance has been shown to account for over 85–90% of the impairment in total body glucose disposal in type 2 diabetic subjects \([6,36]\) (Figure 25.9). There is a significant delay (20–30 min) in the muscle’s response to insulin. However, even if the insulin clamp is extended for an additional hour in diabetic subjects to account for the delay in onset of insulin action, the rate of insulin-stimulated glucose disposal remains 50% less than in control subjects. A similar defect in insulin-stimulated muscle glucose uptake in type 2 diabetic subjects has been demonstrated using the limb catheterization technique \([36,200–203]\).

In type 2 diabetic subjects multiple intramyocellular defects in insulin action have been demonstrated (revised in \([1–3,42]\)), including impaired glucose transport and phosphorylation \([36,203–206]\), reduced glycogen synthesis \([205–208]\), and decreased glucose oxidation \([39,209]\). However, more proximal defects in the insulin signal transduction system play a paramount role in the muscle insulin resistance \([42,210,211]\).)

**Insulin signal transduction.** For insulin to work, it must first bind to and then activate the insulin receptor by phosphorylating key tyrosine residues on the \(\beta\) chain \([42,211–214]\) (Figure 25.10). This results in the translocation of insulin receptor substrate (IRS)-1 to the plasma membrane, where it interacts with the insulin receptor and also undergoes tyrosine phosphorylation. This leads to the activation of PI-3 kinase and Akt, resulting in glucose transport into the cell, activation of nitric oxide synthase with arterial vasodilation \([215–217]\), and stimulation of multiple intracellular metabolic processes.

Studies by DeFronzo and colleagues were the first to demonstrate in humans that the ability of insulin to tyrosine phosphorylate IRS-1 was severely impaired in lean type
2 diabetic individuals [42,210,211,218], in obese normal glucose-tolerant individuals [210], and in the insulin-resistant, normal glucose-tolerant offspring of two type 2 diabetic parents [219] (Figure 25.11). Similar defects have been demonstrated by others in human muscle [37,220–223]. The defect in insulin signaling leads to decreased glucose transport, impaired release of nitric oxide with endothelial dysfunction, and multiple defects in intramyocellular glucose metabolism.

In contrast to the severe defect in IRS-1 activation, the mitogen-activated protein (MAP) kinase pathway, which can be activated by Shc, is normally responsive to insulin [210] (Figure 25.11). The MAP kinase pathway, when stimulated, leads to the activation of a number of intracellular pathways involved in inflammation, cellular proliferation, and atherosclerosis [211,224–226]. The block at the level of IRS-1 impairs glucose transport into the cell and the resultant hyperglycemia stimulates insulin secretion. Because the MAP kinase pathway retains its sensitivity to insulin [210,211,220,226], this causes excessive stimulation of this pathway and activation of multiple intracellular pathways involved in inflammation and atherogenesis. This, in part, may explain the strong association between insulin resistance and atherosclerotic cardiovascular disease in nondiabetic, as well as in type 2 diabetic, individuals [211,227–232]. The only class of oral antidiabetic drugs—the TZDs—that simultaneously augment insulin signaling through IRS-1 and inhibit the MAP kinase pathways is the thiazolidinediones [218].

**Route of glucose administration: oral versus intravenous.** The euglycemic insulin clamp, by maintaining plasma glucose and insulin levels constant, has become the gold standard for quantitating insulin sensitivity. However, the normal route of glucose administration in every day life is via the gastrointestinal tract. Using a double tracer technique (1-14C-glucose orally and 3-3H-glucose intravenously) in combination with hepatic vein catheterization, the disposal of oral versus i.v. glucose has been examined in healthy, normal glucose-tolerant and type 2 diabetic subjects [6,40,190,196,233]. Under basal conditions, with fasting plasma glucose and insulin concentrations of 90 μg dL−1 and 11 mU mL−1, respectively, the splanchnic tissues, which primarily reflect the liver, take up glucose at the rate of 0.5 mg kg−1 per min (Figure 25.12). When insulin is administered intravenously in subjects with NGT to raise the plasma insulin concentration to 1189 μU mL−1 while maintaining euglycemia, no stimulation of hepatic glucose uptake is observed. When insulin is infused with glucose to elevate both glucose and insulin levels, hepatic glucose uptake is increased, but only in proportion to the increase in plasma glucose concentration, despite plasma insulin concentrations in excess of 1000 μU mL−1. In contrast, oral glucose administration augments hepatic glucose uptake 4.5-fold, despite plasma insulin and glucose concentrations that are much lower than with i.v. glucose plus insulin administration (Figure 25.12). If the same oral glucose load is administered to type 2 diabetic individuals, despite higher plasma glucose and insulin concentrations than in nondiabetic subjects, hepatic glucose uptake is reduced by >50%. These results indicate that T2DM individuals lack the gut factor responsible for enhancing hepatic glucose uptake following glucose ingestion.

**Summary: pathogenesis.** In summary, impaired insulin secretion, decreased muscle glucose uptake, increased HGP, and
promoting effects of insulin [234]. In liver, IRS-2 serves as the glucose oxidation, and other acute metabolic and growth the ability of insulin to stimulate muscle glycogen synthesis, domains. Mutation of these specific tyrosines severely impairs recognition sites for proteins containing

goestyrosine phosphorylation in regions containing specific

The cellular events via which insulin initiates its stimulatory effect on glucose metabolism start with binding of the hormone to specific receptors that are present on the cell surface of all insulin target tissues [1–3,211–214]. After insulin has bound to and activated its receptor, “second messengers” are generated and these second messengers activate a cascade of phosphorylation-dephosphorylation reactions that result in stimulation of intracellular glucose metabolism. The first step in glucose utilization involves activation of the glucose transport system, leading to glucose influx into insulin target tissues, primarily muscle. The free glucose, which has entered the cell, subsequently is metabolized by a series of enzymatic steps that are under the control of insulin. Of these, the most important are glucose phosphorylation (catalyzed by hexokinase), glycogen synthase (which controls glycogen synthesis), and phosphofructokinase (PFK) and PDH (which regulate glycolysis and glucose oxidation, respectively).

**Insulin receptor/insulin receptor tyrosine kinase**

The insulin receptor is a glycoprotein which consists of two α-subunits and two β-subunits linked by disulfide bonds [1–3, 211–214] (Figure 25.10). The two α-subunits of the insulin receptor are entirely extracellular and contain the insulin-binding domain. The β-subunits have an extracellular domain, a transmembrane domain, and an intracellular domain that expresses insulin-stimulated kinase activity directed towards its own tyrosine residues. Phosphorylation of the β-subunit, with subsequent activation of insulin receptor tyrosine kinase, represents the first step in the action of insulin on glucose metabolism. Mutagenesis of any of the three major phosphorylation sites (at residues 1158, 1163, and 1162) impairs insulin receptor kinase activity, leading a decrease in the metabolic and growth promoting effects of insulin [234].

**Insulin receptor signal transduction**

Following its activation, insulin receptor tyrosine kinase phosphorylates specific intracellular proteins, of which at least nine have been identified [212,213,235]. In muscle insulin-receptor substrate-1 (IRS-1) serves as the major docking protein that interacts with the insulin receptor tyrosine kinase and undergoes tyrosine phosphorylation in regions containing specific amino acid sequence motifs that, when phosphorylated, serve as recognition sites for proteins containing src-homology 2 (SH2) domains. Mutation of these specific tyrosines severely impairs the ability of insulin to stimulate muscle glycogen synthesis, glucose oxidation, and other acute metabolic and growth promoting effects of insulin [234]. In liver, IRS-2 serves as the primary docking protein that undergoes tyrosine phosphorylation and mediates the effect of insulin on hepatic glucose production, gluconeogenesis, and glycogen formation [236].

In muscle, the phosphorylated tyrosine residues of IRS-1 mediate an association with the 85-kDa regulatory subunit of phosphatidylinositol-3 kinase (PI-3 kinase), leading to activation of the enzyme [1–3,211–214,220,237] (Figure 25.10). PI-3 kinase is comprised of an 85-kDa regulatory subunit and a 110-kDa catalytic subunit. The latter catalyzes the 3-prime phosphorylation of phosphatidylinositol (PI), PI-4-phosphate, and PI-4,5-diphosphate, resulting in the stimulation of glucose transport. Activation of PI-3 kinase by phosphorylated IRS-1 also leads to activation of glycogen synthase, via a process that involves activation of PKB/Akt and subsequent inhibition of kinases, such as glycogen synthase kinase (GSK)-3, and activation of protein phosphatase 1 (PP1). Inhibitors of PI-3 kinase impair glucose transport and block the activation of glycogen synthase and hexokinase (HK)-II expression [211–214,220,237–239]. The action of insulin to increase protein synthesis and inhibit protein degradation also is mediated by PI-3 kinase.

Other proteins with SH2 domains, including the adapter protein Grb2 and Shc, also interact with IRS-1 and become phosphorylated following exposure to insulin [211–214,220]. Grb2 and Shc link IRS-1/IRS-2 to the mitogen-activated protein (MAP) signaling pathway (Figure 25.11), which plays an important role in the generation of transcription factors and promotes cell growth, proliferation, and differentiation [212,220]. Inhibition of the MAP kinase pathway prevents the stimulation of cell growth by insulin but has no effect on the metabolic actions of the hormone [240].

Under anabolic conditions insulin augments glycogen synthesis by simultaneously activating glycogen synthase and inhibiting glycogen phosphorylase [241,242]. The effect of insulin is mediated via the PI-3 kinase pathway which inactivates kinases, such as glycogen synthase kinase-3, and activates phosphatases, particularly protein phosphatase 1 (PP1). PP1 is believed to be the primary regulator of glycogen metabolism. In skeletal muscle, PP1 associates with a specific glycogen-binding regulatory subunit, causing dephosphorylation (activation) of glycogen synthase. PP1 also phosphorylates (inactivates) glycogen phosphorylase. Multiple studies have demonstrated convincingly that inhibitors of PI-3 kinase inhibit glycogen synthase activity and abolish glycogen synthesis [213,243].

**Insulin receptor signal transduction defects in type 2 diabetes**

**Insulin receptor number and affinity**

Both receptor and postreceptor defects contribute to insulin resistance in individuals with T2DM. Some studies have demonstrated a modest 20–30% reduction in insulin binding to monocytes and adipocytes from T2DM patients, but this has not been a consistent finding [1–3,244–247]. The decrease in insulin binding is due to a reduction in the number of
insulin receptors without change in insulin receptor affinity. However, caution should be employed in interpreting these studies, since muscle and liver, not adipocytes, are the major tissues responsible for the regulation of glucose homeostasis in vivo and insulin binding to solubilized receptors obtained from skeletal muscle and liver has been shown to be normal in obese and lean diabetic individuals [245,246,248]. Moreover, a decrease in insulin receptor number cannot be demonstrated in over half of type 2 diabetic subjects, and it has been difficult to demonstrate a correlation between reduced insulin binding and the severity of insulin resistance [249–251]. A variety of defects in insulin receptor internalization and processing have been described in syndromes of severe insulin resistance and diabetes. However, the insulin receptor gene has been sequenced in T2DM patients from diverse ethnic populations and, with very rare exceptions, physiologically significant mutations in the insulin receptor gene have not been observed [252,253]. This excludes a structural gene abnormality in the insulin receptor as a cause of common T2DM.

### Insulin receptor tyrosine kinase activity

Insulin receptor tyrosine kinase activity has been examined in skeletal muscle, adipocytes, and hepatocytes from normal-weight and obese diabetic subjects. Most [1–3,210,246,249,254], but not all [248], investigators have found a reduction in tyrosine kinase activity (Figure 25.11) that cannot be explained by alterations in insulin receptor number or insulin receptor binding affinity. However, restoration of normoglycemia by weight loss has been shown to correct the defect in insulin receptor tyrosine kinase activity [255], suggesting that the defect in tyrosine kinase is acquired secondary to some combination of hyperglycemia, distributed intracellular glucose metabolism, hyperinsulinemia, and insulin resistance—all of which improved after weight loss. Exposure of cultured fibroblasts to high glucose concentration also inhibits insulin receptor tyrosine kinase activity [256]. Since insulin receptor tyrosine kinase activity assays are performed in vitro, the results of these assays could provide misleading information with regard to insulin receptor function in vivo. To circumvent this problem, investigators have employed the euglycemic hyperinsulinemic clamp with muscle biopsies and antiphosphotyrosine immunoblot analysis to provide a “snapshot” of the insulin-stimulated tyrosine phosphorylation state of the receptor in vivo [210]. In insulin-resistant obese nondiabetic and type 2 diabetic subjects a substantial decrease in insulin receptor tyrosine phosphorylation has been demonstrated. However, when insulin-stimulated insulin receptor tyrosine phosphorylation was examined in normal glucose-tolerant, insulin-resistant individuals (offspring of two diabetic parents) at high risk of developing T2DM, a normal increase in tyrosine phosphorylation of the insulin receptor was observed [219]. These findings are consistent with the concept that impaired insulin receptor tyrosine kinase activity in type 2 diabetic patients is acquired secondary to hyperglycemia or some other metabolic disturbance.

### Insulin signaling (IRS-1 and PI-3 kinase) defects

In insulin-resistant obese nondiabetic subjects, the ability of insulin to activate insulin receptor and IRS-1 tyrosine phosphorylation in muscle is modestly reduced, while in T2DM individuals insulin-stimulated insulin receptor and IRS-1 tyrosine phosphorylation are severely impaired [210] (Figure 25.11). Association of the p85 subunit of PI-3 kinase with IRS-1 and activation of PI-3 kinase also are greatly attenuated in obese nondiabetic and type 2 diabetic subjects compared to lean healthy controls [210,220,221] (Figure 25.11). The decrease in insulin-stimulated association of the p85 regulatory subunit of PI-3 kinase with IRS-1 is closely correlated with the reduction in insulin-stimulated muscle glycogen synthase activity and in vivo insulin-stimulated glucose disposal [210]. Impaired regulation of PI-3 kinase gene expression by insulin also has been demonstrated in skeletal muscle and adipose tissue of type 2 diabetic subjects [257]. In animal models of diabetes, an 80–90% decrease in insulin-stimulated IRS-1 phosphorylation and PI-3 kinase activity has been reported [258].

In the insulin-resistant, normal glucose-tolerant offspring of two type 2 diabetic parents, IRS-1 tyrosine phosphorylation and the association of p85 protein/PI-3 kinase activity with IRS-1 are markedly decreased despite normal tyrosine phosphorylation of the insulin receptor; these insulin signaling defects are correlated closely with the severity of insulin resistance, measured with the euglycemic insulin clamp technique [219]. In summary, impaired association of PI-3 kinase with IRS-1 and its subsequent activation are characteristic abnormalities in type 2 diabetics, and these defects are correlated closely with in vivo muscle insulin resistance. A common mutation in the IRS-1 gene (Gly 972 Arg) has been associated with T2DM, insulin resistance, and obesity, but the physiologic significance of this mutation remains to be established [259].

Insulin resistance of the PI-3 kinase signaling pathway contrasts with an intact stimulation of the MAP kinase pathway by insulin in insulin-resistant type 2 diabetic and obese nondiabetic individuals [1–3,209–211,220]. Physiologic hyperinsulinemia increases MEK1 activity and ERK1/2 phosphorylation and activity similarly in lean healthy subjects and in insulin-resistant obese nondiabetic and type 2 diabetic patients. Intact stimulation of the MAP kinase pathway by insulin in the presence of insulin resistance in the PI-3 kinase pathway may play an important role in the development of atherosclerosis [210]. If the metabolic (PI-3 kinase) pathway is impaired, plasma glucose levels rise, resulting in increased insulin secretion and hyperinsulinemia. Because insulin receptor function is normal or only modestly impaired, especially early in the natural history of T2DM, this leads to excessive stimulation of the MAP kinase (mitogenic) pathway in vascular tissues, with resultant proliferation of vascular smooth muscle.
cells, increased collagen formation, and increased production of growth factors and inflammatory cytokines [211,226,260].

Glucose transport
The insulin signal transduction system stimulates glucose transport via a mechanism that involves translocation of a large intracellular pool of glucose transporters (associated with low-density microsomes) to the plasma membrane and their subsequent activation after insertion into the cell membrane [261,262]. There are five major facilitative glucose transporters with distinctive tissue distributions [263,264] (Table 25.2). GLUT4, the insulin regulatable transporter, is found in insulin-sensitive tissues (muscle and adipocytes), has a K_m of ~5 mmol L^-1 which is close to that of the plasma glucose concentration, and is associated with hexokinase (HK)-II [263,264]. In adipocytes and muscle, GLUT4 concentration in the plasma membrane increases markedly after exposure to insulin, and this increase is associated with a reciprocal decline in the intracellular GLUT4 pool. GLUT1 is the predominant glucose transporter in the insulin-independent tissues (brain and erythrocytes), but is also found in muscle and adipocytes. GLUT1 is located primarily in the plasma membrane, where its concentration is unchanged following exposure to insulin. It has a low K_m (~1 mmol L^-1) and is well suited for its function, which is to mediate basal glucose uptake. It is found in association with HKI [265]. GLUT2 is the predominant transporter in liver and pancreatic β cells, where it is found in association with a specific hexokinase, HKIV or glucokinase [266]. GLUT2 has a very high K_m (~15–20 mmol L^-1), which allows the glucose concentration in cells expressing this transporter to rise in direct proportion to the increase in plasma glucose concentration. This unique characteristic allows these cells to function as glucose sensors.

In adipocytes and muscle of type 2 diabetic patients glucose transport activity is severely impaired [220,249,261,262,267,268]. Impaired glucose transport is evident in muscle of diabetic subjects examined using MRI tech-nique, the in vivo dose-response curve for the action of insulin on glucose transport in forearm skeletal muscle has been examined in type 2 diabetic subjects and insulin-stimulated inward muscle glucose transport has been shown to be severely impaired [36,203,271]. Impaired in vivo muscle glucose transport in type 2 diabetes has also been demonstrated using MRI [200] and PET [272]. Since the number of GLUT4 transporters in the muscle of diabetic subjects is normal, impaired GLUT4 translocation and decreased intrinsic activity of the glucose transporter are responsible for the defect in muscle glucose transport. Large populations of type 2 diabetics have been screened for mutations in the GLUT4 gene [273]. Such mutations are very uncommon and, when detected, have been of questionable physiologic significance.

Glucose phosphorylation
Glucose phosphorylation and glucose transport are tightly coupled phenomena [274]. Hexokinase isoenzymes (HK-I–HK-IV) catalyze the first committed step of glucose metabolism, the intracellular conversion of free glucose to glucose-6-phosphate [263–265,275] (G-6-P) (Table 25.2). HK-I, HK-II, and HK-III are single-chain peptides that have a very high affinity for glucose and demonstrate product inhibition by (G-6-P). HK-IV, also called glucokinase, has a lower affinity for glucose and is not inhibited by G-6-P. Glucokinase (HK-IVB) represents the glucose sensor in the β cell, while HK-IVL in the liver plays a central role in the regulation of hepatic glucose metabolism.

In human skeletal muscle, HK-II transcription is regulated by insulin, whereas HK-1 mRNA and protein levels are not affected by insulin [276–278]. In response to physiologic euglycemic hyperinsulinemia of 2–4 h duration, HK-II cytosolic activity, protein content, and mRNA levels increase by 50–200% in healthy nondiabetic subjects and this is associated with the translocation of HK-II from the cytosol to the mitochondria. In forearm muscle, insulin-stimulated glucose transport (measured with the triple tracer technique) is markedly impaired in lean type 2 diabetics [36,203,271]. However, the rate of intracellular glucose phosphorylation is impaired to an even greater extent, resulting in an increase in the free glucose concentration within the intracellular space that is accessible

<table>
<thead>
<tr>
<th>Organ</th>
<th>Glucose transporter</th>
<th>HK</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brain</td>
<td>GLUT1</td>
<td>HK-I</td>
<td>Glucose dependent</td>
</tr>
<tr>
<td>Erythrocyte</td>
<td>GLUT1</td>
<td>HK-I</td>
<td>Glucose dependent</td>
</tr>
<tr>
<td>Adipocyte</td>
<td>GLUT4</td>
<td>HK-II</td>
<td>Insulin dependent</td>
</tr>
<tr>
<td>Muscle</td>
<td>GLUT4</td>
<td>HK-II</td>
<td>Insulin dependent</td>
</tr>
<tr>
<td>Liver</td>
<td>GLUT2</td>
<td>HK-IVL</td>
<td>Glucose sensor</td>
</tr>
<tr>
<td>GK β cell</td>
<td>GLUT2 HK-IVB (glucokinase)</td>
<td>Glucose sensor</td>
<td></td>
</tr>
<tr>
<td>Gut</td>
<td>GLUT3-symporter</td>
<td>—</td>
<td>Sodium dependent</td>
</tr>
<tr>
<td>Kidney</td>
<td>GLUT3-symporter</td>
<td>—</td>
<td>Sodium dependent</td>
</tr>
</tbody>
</table>

to glucose. These observations indicate that in type 2 diabetic individuals, while both glucose transport and glucose phosphorylation are severely resistant to the action of insulin, impaired glucose phosphorylation (HK-II) appears to be the rate-limiting step for insulin action. Studies using $^{31}$P-NMR in combination with 1-14C-glucose have also demonstrated that both insulin-stimulated muscle glucose transport and glucose phosphorylation are impaired in type 2 diabetic subjects, but results from this study suggest that the defect in transport exceeds the defect in phosphorylation [205]. Because of methodologic differences, the results of the triple tracer [36,203,271] and MRI [205] studies cannot be reconciled at present. Nonetheless, these studies are consistent in demonstrating that abnormalities in both muscle glucose phosphorylation and glucose transport are well established early in the natural history of T2DM and cannot be explained by glucose toxicity.

In healthy nondiabetic subjects, a physiologic increase in the plasma insulin concentration for as little as 2–4 hours increases muscle HK-II activity, gene transcription, and translation [276]. In lean type 2 diabetics the ability of insulin to augment HK-II activity and mRNA levels is markedly reduced compared to controls [277]. Decreased basal muscle HK-II activity and mRNA levels and impaired insulin-stimulated HK-II activity in type 2 diabetic subjects have been reported by other investigators [278,279]. A decrease in insulin-stimulated muscle HK-II activity has also been described in subjects with IGT [280]. Several groups have looked for point mutations in the HKII gene in individuals with T2DM and, although several nucleotide substitutions have been found, none are close to the glucose and ATP binding sites and none have been associated with insulin resistance [280–282]. Thus, an abnormality in the HKII gene is unlikely to explain the inherited insulin resistance in common variety T2DM mellitus.

**Glycogen synthesis**

Following phosphorylation by HK-II, glucose either can be converted to glycogen or enter the glycolytic pathway. Of the glucose that enters the glycolytic pathway, ∼90% is oxidized and the remaining 10% is released as lactate. At low physiologic plasma insulin concentrations, glycogen synthesis and glucose oxidation contribute approximately equally to glucose disposal. However, with increasing plasma insulin concentrations, glycogen synthesis predominates [1–3,283]. Impaired insulin-stimulated glycogen synthesis is a characteristic finding in all insulin-resistant states including obesity, IGT, diabetes, and diabesity in all ethnic groups and accounts for the majority of the defect in insulin-mediated whole body glucose disposal [1–3,12,162,284–287]. Impaired glycogen synthesis also has been documented in the normal glucose-tolerant offspring of two diabetic parents, in the first-degree relatives of type 2 diabetic individuals, and in the normoglycemic twin of a monozygotic twin pair in which the other twin has T2DM [162,288,289].

Glycogen synthase is the key insulin-regulated enzyme which controls the rate of muscle glycogen synthesis [241, 243,278,289–291]. Insulin activates glycogen synthase by stimulating a cascade of phosphorylation-dephosphorylation reactions, which ultimately lead to the activation of PP1 (also called glycogen synthase phosphatase). The regulatory subunit of PP1 has two serine phosphorylation sites. Phosphorylation of site 2 by cAMP-dependent kinase (PKA) inactivates PP1, while phosphorylation of site 1 by insulin activates PP1, leading to the stimulation of glycogen synthase. Phosphorylation of site 1 of PP1 by insulin in muscle is catalyzed by insulin-stimulated protein kinase 1 (ISPK-1). Because of their central role in muscle glycogen formation, the three enzymes—glycogen synthase, PP1, ISPK-1—have been extensively studied in the individuals with T2DM.

Glycogen synthase exists in an active (dephosphorylated) and an inactive (phosphorylated) form [241–243]. Under basal conditions, total glycogen synthase activity in type 2 diabetic subjects is reduced and the ability of insulin to activate glycogen synthase is severely impaired [210,292–294]. The ability of insulin to stimulate glycogen synthase is also diminished in the normal glucose-tolerant, insulin-resistant relatives of type 2 diabetic individuals [295]. In insulin-resistant nondiabetic and diabetic Pima Indians activation of muscle PP1 (glycogen synthase phosphatase) by insulin is severely reduced [296]. Since PP1 dephosphorylates glycogen synthase, leading to its activation, the defect in PP1 plays an important role in the muscle insulin resistance of T2DM.

The effect of insulin on glycogen synthase gene transcription and translation in vivo has been studied extensively. Most studies have demonstrated that insulin does not increase glycogen synthase mRNA or protein expression in human muscle [276,297,298]. However, glycogen synthase mRNA and protein levels are decreased in muscle of type 2 diabetic patients, explaining in part the decreased glycogen synthase activity [298,299]. The major abnormality in glycogen synthase regulation in T2DM is its lack of dephosphorylation and activation by insulin, as a result of insulin receptor signaling abnormalities (see previous discussion).

The glycogen synthase gene has been the subject of intensive investigation, and DNA sequencing has revealed either no mutations or rare nucleotide substitutions that cannot explain the defect in insulin-stimulated glycogen synthase activity [300–302]. The genes encoding the catalytic subunits of PP1 and ISPK-1 have been examined in Pima Indians and Danes with T2DM [303,304]. Several silent nucleotide substitutions were found in the PP1 and ISPK-1 genes in the Danish population, but the mRNA levels of both genes were normal in skeletal muscle. No structural gene abnormalities in the catalytic subunit of PP1 were detected in Pima Indians. Thus, neither mutations in the PP1 and ISPK-1 genes nor abnormalities in their translation can explain the impaired enzymatic activities of glycogen synthase and PP1 that have been observed in vivo. Similarly, there is no evidence that an alteration in glycogen
phosphorylase plays any role in the abnormality in glycogen formation in T2DM [305].

In summary, glycogen synthase activity is severely impaired in type 2 diabetic individuals, but the molecular etiology of the defect remains to be determined.

Glycolysis/glucose oxidation

Glucose oxidation accounts for ~90% of total glycolytic flux, while anaerobic glycolysis accounts for the other 10%. Two enzymes, phosphofructokinase (PFK) and pyruvate dehydrogenase (PDH), play pivotal roles in the regulation of glycolysis and glucose oxidation, respectively. In type 2 diabetic individuals the glycolytic/glucose oxidative pathway has been shown to be impaired [286]. Although one study has suggested that PFK activity is modestly reduced in muscle biopsies from type 2 diabetic subjects [306], most evidence indicates that the activity of PFK is normal [293,298]. Insulin has no effect on muscle PFK activity, mRNA levels, or protein content in either nondiabetic or diabetic individuals [298]. PDH is a key insulin-regulated enzyme whose activity in muscle is acutely stimulated by insulin [307]. In type 2 diabetic patients, insulin-stimulated PDH activity is decreased in human adipocytes and in skeletal muscle [307,308].

Obesity and T2DM are associated with accelerated FFA turnover and oxidation [1–3,12,309], which would be expected, according to the Randle cycle [310], to inhibit PDH activity and consequently glucose oxidation. Therefore, it is likely that the observed defects in glucose oxidation and PDH activity are acquired secondary to increased FFA oxidation and feedback inhibition of PDH by elevated intracellular levels of acetyl-CoA and reduced availability of NAD. Consistent with this scenario, the rates of basal and insulin-stimulated glucose oxidation are not reduced in the normal glucose-tolerant offspring of two diabetic parents and in the first-degree relatives of type 2 diabetic subjects, while it is decreased in overtly diabetic subjects.

Summary

In summary, postbinding defects in insulin action primarily are responsible for the insulin resistance in T2DM. Diminished insulin binding, when present, is modest and secondary to downregulation of the insulin receptor by chronic hyperinsulinemia. In type 2 diabetic patients with overt fasting hyperglycemia, a number of postbinding defects have been demonstrated, including reduced insulin receptor tyrosine kinase activity, insulin signal transduction abnormalities, decreased glucose transport, diminished glucose phosphorylation, and impaired glycogen synthase activity. The glycolytic/glucose oxidative pathway is largely intact and, when defects are observed, they appear to be acquired secondary to enhanced FFA/lipid oxidation. From the quantitative standpoint, impaired glycolysis synthesis represents the major pathway responsible for the insulin resistance in T2DM, and is present long before the onset of overt diabetes, that is, in normal glucose-tolerant, insulin-resistant prediabetic subjects and in individuals with IGT. The impairment in glycogen synthase activation appears to result from a defect in the ability of insulin to phosphorylate IRS-1, causing a reduced association of the p85 subunit of PI-3 kinase with IRS-1 and decreased activation of the enzyme PI-3 kinase.

The adipocyte, FFA metabolism, and lipotoxicity

Considerable evidence implicates deranged adipocyte metabolism and altered fat topography in the pathogenesis of glucose intolerance in T2DM [1–3,39,110,197,311–314]: (i) fat cells are resistant to insulin’s antilipolytic effect, leading to day-long elevation in the plasma FFA concentration [1–3,39,110,208,311–315]; (ii) chronically increased plasma FFA levels stimulate gluconeogenesis [316–318], induce hepatic/muscle insulin resistance [319–321], and impair insulin secretion [110,322]. These FFA-induced disturbances are referred to as lipotoxicity; (iii) dysfunctional fat cells produce excessive amounts of insulin resistance–inducing, inflammatory, and atherosclerotic-provoking adipocytokines and fail to secrete normal amounts of insulin-sensitizing adipocytokines such as adiponectin [311,312]; (iv) enlarged fat cells are insulin resistant and have diminished capacity to store fat [323,324]. When adipocyte storage capacity is exceeded, lipid “overflows” into muscle, liver, and β cells, causing muscle/hepatic insulin resistance and impaired insulin secretion (reviewed in [311] and [312]). This represents another form of lipotoxicity. Lipid can also overflow into arterial vascular smooth cells, leading to the acceleration of atherosclerosis.

Using 14C-palmitate in combination with the insulin clamp technique [39], the antilipolytic effect of insulin has been shown to be markedly impaired in lean type 2 diabetic subjects, as well as in obese nondiabetic subjects [208]. In both type 2 diabetic (Figure 25.13) and obese nondiabetic subjects, the ability of insulin to suppress the plasma FFA concentration and inhibit FFA turnover is impaired compared with lean normal glucose-tolerant subjects at all plasma insulin-concentrations spanning the physiologic and pharmacologic range.

Many investigators [317,318,321,325] have demonstrated that a physiologic elevation in the plasma FFA concentration stimulates HGP and impairs insulin-stimulated glucose uptake in liver [326] and muscle [318–321,325–330]. Chronically elevated plasma FFA levels also inhibit insulin secretion [110,322], especially in genetically prone individuals.

According to the Randle [310,331] cycle of substrate competition, elevated FFA oxidation in muscle reciprocally impairs glucose oxidation. Although clearly there is substrate competition between FFA and glucose with respect to oxidative metabolism [323,333], FFAs have been shown to have independent effects to inhibit glycogen synthase [334,335] and both glucose transport and glucose phosphorylation [328,336].
Further, physiologic elevation in the plasma FFA concentration for as little as 4 hours markedly impairs insulin signal transduction and inhibits insulin-mediated glucose disposal by 30–35% in healthy lean normal glucose-tolerant subjects [337]. The elevation in plasma FFA concentration caused a dose-response inhibition of muscle insulin receptor tyrosine phosphorylation, IRS-1 tyrosine phosphorylation, PI-3 kinase activity, and Akt serine phosphorylation (Figure 25.14). Conversely, reduction in the plasma FFA concentration with acipimox in T2DM individuals enhances insulin sensitivity by ~30% in association with an increase in insulin signaling, glycogen synthesis, and glucose oxidation [338,339].

After fatty acids enter the cell, they can be converted to triglycerides, which are inert, or to toxic lipid metabolites such as fatty acyl CoAs, diacylglycerol, and ceramide [4]. Using magnetic resonance spectroscopy, the intramyocellular triglyceride content has been shown to be increased in type 2 diabetic subjects [338,340]. Fatty acyl CoAs, which are known to inhibit insulin signaling [341,342], are also significantly increased in muscle in diabetic subjects [338,339].

Peroxisome proliferator-activated γ coactivator-1 (PGC-1) is the master regulator of mitochondrial biogenesis [344] and augments the expression of multiple genes involved in mitochondrial oxidative phosphorylation [345–347]. In individuals with T2DM and in the normal glucose-tolerant, insulin-resistant offspring of two diabetic parents the expression of PGC-1, nuclear receptor factor-1, and multiple other genes involved in oxidative phosphorylation are markedly reduced in muscle and are strongly correlated with the defect in glucose oxidation and whole body (muscle) insulin resistance [211,343,348]. The reduced expression and activity of these key mitochondrial genes in the NGT offspring strongly suggests a genetic etiology for the mitochondrial dysfunction. However, there also is evidence that the mitochondrial defect is acquired, at least in part [349–351]. Treatment of diabetic patients with pioglitazone markedly improves insulin sensitivity in association with a reduction in intramyocellular lipid and fatty acyl CoA concentrations. The decrement in muscle fatty acyl CoA content is closely related to the improvement in insulin-stimulated muscle glucose disposal [340,343]. Reduced intramyocellular fatty acyl CoA content with acipimox in T2DM individuals enhances insulin sensitivity by ~30% in association with an increase in insulin signaling, glycogen synthesis, and glucose oxidation [338,339].

Figure 25.13 Dose-response relationship between the plasma insulin concentration and suppression of plasma FFA concentration, FFA turnover, and FFA oxidation in type 2 diabetic (solid circles) and normal glucose-tolerant (open circles) subjects. Source: Groop L, et al. Journal of Clinical Investigation 1989;84:205–215.

these findings provide strong support for lipotoxicity and adipocyte insulin resistance in the pathogenesis of T2DM.

**Alpha cell and glucagon**

It long has been known that the basal plasma glucagon concentration is elevated in type 2 diabetic individuals [184–186, 357,358]. The important contribution of elevated fasting plasma glucagon levels to the increased basal rate of HGP in type 2 diabetic individuals was provided by Baron et al. [187]. Compared with control subjects, diabetic individuals had a markedly elevated rate of basal HGP, which correlated closely with the increase in fasting plasma glucagon concentration. Somatostatin infusion reduced the plasma glucagon concentration by 44% in association with a 58% decrease in basal HGP (Figure 25.15). These results conclusively demonstrate the pivotal role of hyperglucagonemia in the pathogenesis of fasting hyperglycemia in T2DM. There also is evidence that the liver is hypersensitive to the stimulatory effect of glucagon in hepatic gluconeogenesis [185].

**The kidney: increased glucose reabsorption**

The kidney filters \( \sim 162 \) g \( (\text{glomerular filtration rate} = 180 \text{L day}^{-1} \times \text{fasting plasma glucose} = 900 \text{mg L}^{-1}) \) of glucose every day. Ninety percent of the filtered glucose is reabsorbed by the high capacity SGLT2 transporter in the convoluted segment of the proximal tubule, and the remaining 10% of the filtered glucose is reabsorbed by the SGLT1 transporter in the straight segment of the descending proximal tubule [362]. The result is that no glucose appears in the urine.

In normal glucose-tolerant subjects, plasma glucagon levels decline following a meal and the decrease in portal vein glucagon concentration contributes to the suppression of HGP [359]. In contrast, following ingestion of a mixed meal in T2DM patients there is a paradoxical rise in plasma glucagon concentration which antagonizes the decline in HGP, resulting in postprandial hyperglycemia [360,361]. Thus, deranged glucagon secretion by the pancreatic α cell contributes to both fasting and postprandial hyperglycemia in T2DM patients.

![Figure 25.15](image1)  
**Figure 25.15** Effect of somatostatin (SRIF) infusion with basal insulin replacement on basal (fasting) hepatic glucose production (HGP) (left) and plasma glucagon concentration (right) in normal glucose-tolerant control (CON) and type 2 diabetic (DIAB) subjects. Normalization of the plasma glucagon concentration reduced HGP by 58% to values observed in CON subjects. Source: Baron AD, et al. *Diabetes* 1987;36:274–283.

![Figure 25.16](image2)  
**Figure 25.16** SGLT2 transporter mRNA (left) and protein (middle) and glucose transport activity (α-methyl-D-glucopyranoside) (right) are increased in cultured renal proximal tubular epithelial cells of individuals with type 2 diabetes (T2DM) versus nondiabetic subjects (CON). Source: Rahmoune H, et al. *Diabetes* 2005;54:3427–3434.
The brain

The brain, along with its seven companions, forms the eighth component of the Ominous Octet (Figure 25.17). The current epidemic of diabetes is being driven by the epidemic of obesity [369]. Porte and colleagues [370–373] were amongst the first to demonstrate that, in rodents, insulin was a powerful appetite suppressant. Obese individuals, both diabetic and nondiabetic, have moderate-to-severe insulin resistance with compensatory hyperinsulinemia. Nonetheless, food intake is increased in obese subjects despite the presence of hyperinsulinemia which should suppress the appetite. Therefore, one could postulate that the insulin resistance in peripheral tissues also extends to the brain.

Using functional magnetic resonance imaging (MRI), the cerebral response to an ingested glucose load has been studied [374]. After glucose ingestion, two hypothalamic areas with consistent inhibition have been noted: the lower posterior hypothalamus, which contains the ventromedial nuclei, and the upper posterior hypothalamus, which contains the paraventricular nuclei. In both of these hypothalamic areas, which are key centers for appetite regulation, the magnitude of the inhibitory response following glucose ingestion was reduced in obese, insulin-resistant, normal glucose-tolerant subjects, and there was a delay in the time taken to reach the maximum inhibitory response, even though the plasma insulin response was markedly increased in the obese group. Whether the impaired functional MRI response in obese subjects contributes to or is a consequence of the insulin resistance and weight gain remains to be determined. Nonetheless, these results suggest that the brain, like other organs (liver, muscle, and fat) in the body, are resistant to insulin. Studies by Obici et al. [375,376] and others [377] in rodents have also provided evidence for cerebral insulin resistance leading to increased HGP and reduced muscle glucose uptake.

Implications for therapy

The preceding review of the pathophysiology of T2DM has important therapeutic implications (Table 25.1). First, effective treatment of T2DM will require multiple drugs used in combination to correct the multiple pathophysiologic defects. Second, the treatment should be based upon reversal of known pathogenic abnormalities and NOT simply on the reduction in HbA1c. Third, therapy must be started early in the natural history of T2DM, if progressive β-cell failure is to be prevented. The treatment of T2DM is discussed in detail in Chapters 42–46.

References


14 Cherrington AD: Control of glucose uptake and release by the liver. *Diabetes* 1999;48:1198 – 1214.


31 Weyer C, Tataranni PA, Bogardus C, Pratley RE: Insulin resistance and insulin secretory dysfunction are independent predictors of worsening of glucose tolerance during each stage of type 2 diabetes development. *Diabetes Care* 2001;24:89 – 94.


119 Goldfine AB, Kulkarni RN: Modulation of β-cell function: a translational journal from the bench to the bedside. *Diabetes, Obesity and Metabolism* 2012;14(Suppl 3):152–160.


208 Shulman GI, Rothman DL, Jue T, et al.: Quantitation of muscle glycogen synthesis in normal subjects and subjects with...


217 Montagnani M, Chen H, Barr VA, Quon MJ: Insulin-stimulated activation of eNOS is independent of Ca2+ but requires phosphorylation by Akt at Ser(1179). The Journal of Biological Chemistry 2001;276:30392–30398.


303 Bjorbaek C, Fik TA, Echward SM, et al.: Cloning of human insulin-stimulated protein kinase (ISPK-1) gene and analysis of coding regions and mRNA levels of the ISPK-1 and the protein...


336 Pendergrass M, Nucci G, DeFronzo R: In vivo glucose transport (GT) and phosphorylation (GP) in skeletal muscle are impaired by elevation of plasma FFA. *Diabetes* 1998;47(Suppl 1):A65.


CHAPTER 26

The genetics of type 2 diabetes

Emma Ahlqvist1 and Leif Groop1,2
1Department of Clinical Sciences, Lund University, University Hospital Skåne, Malmö, Sweden
2Finnish Institute for Molecular Medicine (FIMM), University of Helsinki, Helsinki, Finland

Key points

• The risk of developing type 2 diabetes depends on both environ-
  mental and genetic factors but their relative importance in the gen-
  eral population is difficult to determine.
• The genetic architecture of T2DM is complex and mostly unknown
  but involves a large number of genetic risk variants of which more
  than 75 have been identified.
• Most loci so far have been identified in genome-wide association
  studies. The risk variants are common in the population, have only
  a small effect on disease risk and together only explain a fraction
  of the heritability of the disease.
• Recent advances in sequencing technology now allows whole
  genome/exome sequencing in larger cohorts which may enable
  identification of rarer variants with stronger effects.
• The missing heritability not explained by the hitherto identified loci
  could be due to yet unidentified common variants, rare variants,
  gene–gene or gene–environment interactions or epigenetic
  effects.

The diabetic spectrum

Diabetes is a group of metabolic diseases characterized by
hyperglycemia resulting from defects in insulin secretion,
insulin action, or both. The chronic hyperglycemia of diabetes
is associated with long-term damage, dysfunction, and failure
of different organs, especially the eyes, kidneys, nerves, heart,
and blood vessels.

There are several types of diabetes with differing disease
mechanisms and clinical characteristics (Figure 26.1). Type
2 diabetes (T2DM) is the most common form, constituting
80–90% of all diabetics. T2DM develops when pancreatic
β-cells can no longer increase their insulin secretion enough
to compensate for increasing insulin resistance imposed by
increasing obesity. There is no formal definition of T2DM;
instead diabetes patients who do not fulfill the criteria of any
other form of diabetes (see later) are considered to have T2DM.
While T2DM onset is usually after the age of 35, it is not
uncommon that T2DM is diagnosed already in adolescents in
high-risk regions such as Asia, the Middle East, and USA.

Type 1 diabetes (T1DM) affects ~8% of diabetes patients and
is due to autoimmune destruction of pancreatic β-cells and is
characterized by complete lack of insulin secretion, presence of
autoantibodies to glutamic acid decarboxylase (GAD) and, usu-
ally, onset of disease before age 35. LADA (latent autoimmune
diabetes in adults) accounts for about 7% of all diabetic patients
in Europe and is usually defined as GAD antibody-positive dia-
betes with onset greater than 35 years of age [1]. These disease
types could be divided further into subtypes possibly represent-
ing different disease mechanisms. In addition there are less com-
mon diabetes types including MODY (maturity onset diabetes
of the young), representing a monogenic form of diabetes with
well-defined mutations in more than six different genes, as well
as neonatal and secondary diabetes. These forms of diabetes rep-
resent a range of genetic etiologies from the monogenic MODY
variants to T2DM, which is a highly complex multigenic disease
with a very strong environmental component.

The diabetes epidemic

Worldwide prevalence figures estimate that there were 371
million diabetic patients in 2012 and more than 500 million are
expected in 2030 (www.diabetesatlas.org). In Europe, ~ 8% of
the population suffer from diabetes and up to 90% of these have
T2DM, making this the fastest increasing disease in Europe
and worldwide. The T2DM epidemic can largely be ascribed
to the worldwide increase in obesity during the last 30 years,
for example more than 60% of individuals older than 15 in the
UK and US are overweight (BMI > 25). This in turn has been
ascribed to a collision between genes and the environment.
In a situation of affluence most people tend to overconsume

© 2015 John Wiley & Sons, Ltd. Published 2015 by John Wiley & Sons, Ltd.
Heritability of T2DM

It has long been clear that T2DM clusters in families and it is well established that the risk of developing T2DM depends on both genetic and environmental factors. However, the exact proportion of the two, that is, the heritability, is difficult to determine and heritability estimates therefore vary in the range of 25–80% between studies. Twin-studies have estimated the genetic component by comparing disease concordance in monzygotic twins with concordance in dizygotic twins. In these studies probandwise concordance rates (number of affected twins having a diabetic co-twin) for monzygotic twins vary between 34 and 100% [3–6].

The relative risk for first-degree relatives, that is, the risk of developing T2DM if you have an affected parent or sibling compared to the general population, is approximately 3, and ~6 if both parents are affected [7]. However, this statistic also varies depending on the cohort and population studied.

There are also large differences between ethnic groups that seem to depend on genetic factors. The prevalence of T2DM varies widely among populations, from a few percent among Caucasians in Europe to as high as 50% among Pima Indians in Arizona [8]. While part of the observed ethnic variability could be attributed to environmental and cultural factors some of the variation seems to depend on genetic differences.

One limitation of both twin and family studies is that they have a risk of overestimating heritability due to sampling and ascertainment errors. Concordant twins are more likely to participate in a study and the proportion of individuals with undiagnosed diabetes may differ compared to the general population. Also, the families recruited to genetic studies may represent a subset of highly heritable T2DM whereas an outbred population could have more sporadic or environmentally induced cases.

Another limitation is that family studies often cannot distinguish between the effect of genetic inheritance and the effects of a shared environment and the gene–environment correlation that comes from nongenetic inheritance. This problem could be especially large in T2DM since poor eating and exercise habits are often part of the social inheritance and strongly affect diabetes risk.

Another interesting fact is that the risk of T2DM is higher in individuals who have an affected mother, compared to an affected father [9]. There are many possible explanations for this including a role for the intrauterine environment in programming events later in life. Intrauterine effects can also affect heritability estimates because monozygotic twins are often monochorionic, which results in growth retardation compared to dizygotic twins, and low birth weight is associated with increased risk of T2DM later in life.

In spite of these reservations, there is no doubt that the risk of T2DM is partly determined by genetic factors, many of which have already been identified, and while each identified variant explains only a very small proportion of the risk of T2DM in the human population they all contribute to our understanding of disease pathogenesis. One should also keep in mind that the variance explained by a risk allele in a population is not necessarily an indicator of its importance in specific patients, nor is it proportional to the affected pathway’s importance or potential as a therapeutic target.

Identification of T2DM affecting genetic variants

Linkage studies

The methods used to map disease-causing variation have evolved rapidly in the last decades thanks to technical advances in genotyping methods (Figure 26.2). Originally, disease-causing loci were identified primarily by linkage analysis, utilizing the long stretches of linkage in affected families. By genotyping 400–500 genetic markers, disease loci can be mapped on a genome-wide level without any prior hypothesis.
about which genes are involved. Finding that affected family members share a certain marker that is identical by descent (i.e., identical because it was inherited from the same parent) more often than expected by chance is evidence that a disease-causing variant is in linkage with that marker. This strategy has been very successful in mapping genetic diseases like MODY that have a strong penetrance and a known mode of inheritance. The first MODY locus, MODY-2, was mapped to the glucokinase (GK) locus on chromosome 7 in 1992 [10]. It was soon followed by the HNF1A (MODY-3) and HNF4A (MODY-1) loci [11,12]. The HNF homeobox genes are transcription factors expressed in liver and β-cells that are necessary for proper regulation of insulin secretion. Common variants in both these loci have since been shown to affect risk of T2DM [13].

Linkage analysis has, however, been less useful for identifying genes causing complex diseases. Even though great efforts have been put into linkage studies of T2DM, only two genes can be claimed to have been identified using this strategy. The first T2DM gene mapped by linkage analysis was CAPN10 on chromosome 10, encoding calpain 10, a cysteine protease with largely unknown functions in glucose metabolism [14]. Despite a number of negative replication studies, several meta-analyses have shown consistent association with T2DM [15,16]. Nevertheless, none of the large genome-wide association studies (GWAS) have identified CAPN10 as being associated with T2DM.

The second locus was first mapped to a 10.5 Mb region on chromosome 10q and was later fine-mapped in the Icelandic population to an intronic variant in the TCF7L2 gene contributing to, but not fully explaining, the original linkage [17–19]. This association has since been confirmed in African, Asian, and European populations making it the best replicated genetic association with T2DM to date, conferring a relative risk of ∼1.4 [20].

Association studies on candidate genes
A popular hypothesis about the genetic architecture of complex diseases such as T2DM, the common disease/common variant hypothesis, suggests that common disorders are caused by aggregation of common risk alleles [21,22]. This model has been the basis of a revolution in complex genetics by stimulating the development of tools for genetic association studies. Association studies utilize the very short LD stretches in unrelated individuals to map risk variants in populations instead of families. The advantages of association lie in the larger number of individuals that can be collected for each study as well as the much higher resolution of the mapping. One disadvantage is the huge number of markers needed to perform mapping on a genome-wide level and association studies were therefore originally performed on small regions known to harbor candidate genes.

The first gene reproducibly associated with T2DM was PPARγ, encoding the nuclear receptor PPAR-γ [23]. The PPAR-γ receptor is a molecular target for thiazolidinedione compounds, a class of insulin-sensitizing drugs used to treat T2DM, making it a very compelling candidate gene. The transcript expressed in adipose tissue has an extra exon B and a substitution of a proline for alanine at position 12 of this protein, which is seen in about 15% of the European population. This variant, has been shown to be associated with
increased transcriptional activity, increased insulin sensitivity, and protection against T2DM [23].

The ADRA2A (adrenergic receptor alpha 2) locus was recently identified as a T2DM risk locus after first having been positionally mapped in congenic GK rats where it was associated with impaired insulin granule docking and reduced β-cell exocytosis [24]. Human carriers of the ADRA2A risk variant (rs553668) have reduced fasting insulin and decreased insulin secretion as a consequence of increased expression of the ADRA2 receptor in pancreatic islets. It is well known that epinephrine excess can suppress insulin secretion and cause diabetes.

Genome-wide association studies (GWAS)

Rapid improvement in high throughput technology for SNP genotyping, allowing simultaneous genotyping of hundreds of thousands of SNPs, has opened new possibilities for association studies. The HapMap project provided another important tool, showing that genotyping of approximately 500,000 SNPs is enough to cover about 75% of the common variants (minor allele frequency >5%) in the genome. Several GWAS for diabetes were published in 2007, coined “Breakthrough of the Year” by Science magazine. The first was a GWAS on early-onset T2DM that was published in February 2007 reporting two new diabetes loci: HHEX and SLC30A8 [25]. A few months later it was followed by four European case-control studies including patients with classical T2DM [26–29]. Three of the studies shared results prior to publication and only considered positive results that were replicated in all studies. This resulted in the identification of two new diabetes loci, CDKN2A/B and IGF2BP2, in addition to confirming previously known loci. CDKAL1 (CDK5 regulatory subunit associated protein 1-like 1) was independently identified as a new T2DM locus in all four studies.

About the same time FTO was identified as a major susceptibility locus for obesity, and therefore indirectly also for T2DM [30,31]. Since the effect of FTO on diabetes is through obesity it was not detected in the GWAS studies of T2DM that were matched on BMI.

The first GWA studies on T2DM in non-European populations were published in 2008 using a multistage approach [32,33]. Both studies identified KCNQI, which encodes the pore-forming alpha subunit of the IK1,K+ channel (or voltage-gated potassium channel), and showed that the gene is expressed in the pancreas. Later a SNP in KCNQ1 was shown to have distorted parent-of-origin transmission in an Icelandic study, the risk allele having a much stronger effect when transmitted from the mother than from the father [34].

The first wave of GWAS was followed by a second wave combining existing or new GWAS into meta-analyses of >50,000 individuals [13,35]. A prerequisite for this was that many research groups could work together in consortia like DIAGRAM (Diabetes Genetics Replication and Meta-analysis) consortium and MAGIC (Meta-Analyses of Glucose- and Insulin-related traits consortium). These very large studies have identified several new loci and confirmed the effect of many previously identified. This has also been facilitated by the 1000 Genomes Project, another international collaboration to produce a public catalog of human genetic variation, including SNPs and structural variants, which now allows imputation and analysis of more than 30 million common and low-frequency variants.

Recently, a number of GWAS and meta-analysis studies have also been performed in non-European cohorts, adding several new loci to the list of genome-wide significant associations [35–45]. Interestingly it seems like most associations found in one ethnic group also show some evidence of association in populations with other ethnicities. In total, GWAS have provided more than 75 loci for T2DM (Table 26.1) as well as numerous loci for glucose- or insulin-related traits and more are likely to come.

High density mapping

GWAS do not inevitably lead to identification of a gene or genes in a given locus associated with disease. The most strongly associated SNPs are often only markers for the functional variant responsible for the observed genetic effect and most associated regions harbor several genes. Therefore, additional fine mapping of the loci in even larger sample sets is often necessary. To do this cost-efficiently a Cardio-MetaboChip has been developed for metabolic/cardiovascular gene mapping. This custom-design Illumina Infinium genotyping chip contains ~200,000 polymorphisms selected to cover association signals from a wide range of metabolic disorders (T2DM, lipid disorders, obesity, and cardiovascular disease), was designed to perform both deep replication of major disease signals and fine mapping of established loci. Meta-analysis of previous GWAS with an additional 22,669 T2DM cases and 58,119 controls genotyped using the Cardio-Metabochip has recently added another eight new loci associated with T2DM in the European population [46].

Next-generation sequencing will provide even denser coverage of genetic variation. Scientists in Iceland recently identified four new rare and low-frequency variants (minor allele frequency 0–5%) associated with T2DM through whole-genome sequencing of 2630 Icelanders and imputation of genotypes in more than 290,000 closely related individuals [47]. The GoT2D consortium aims to map lower frequency variants via low-coverage whole-genome sequencing, deep exome sequencing, and next-generation 2.5M SNP chip genotyping of 1325 cases and 1325 controls selected in the phenotype extremes of T2DM. A custom-designed Exome Chip containing rare variants that have been seen at least three times in different studies has also been designed and is currently used to try to identify rare variants associated with T2DM.

Functions of associated genes

Most identified diabetes loci have not been mechanistically tied to the disease. While loci are commonly referred to by the
Table 26.1 Genetic loci associated with risk of T2DM

<table>
<thead>
<tr>
<th>Variants</th>
<th>Candidate gene</th>
<th>Chr</th>
<th>OR</th>
<th>R/NR</th>
<th>RAF</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs4607103</td>
<td>ADAMTS9</td>
<td>3</td>
<td>1.09</td>
<td>C/T</td>
<td>0.81</td>
<td>[46,63]</td>
</tr>
<tr>
<td>rs6705735</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2877716</td>
<td>ADCYS</td>
<td>3</td>
<td>1.12</td>
<td>C/T</td>
<td>0.75</td>
<td>[87]</td>
</tr>
<tr>
<td>rs515071</td>
<td>ANK1</td>
<td>8</td>
<td>1.18</td>
<td>G/A</td>
<td>0.81</td>
<td>[37,46]</td>
</tr>
<tr>
<td>rs516946</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs459193</td>
<td>ANKRDS5</td>
<td>5</td>
<td>1.08</td>
<td>G/A</td>
<td>0.78</td>
<td>[46]</td>
</tr>
<tr>
<td>rs2028299</td>
<td>APJS2</td>
<td>5</td>
<td>1.10</td>
<td>C/A</td>
<td>0.73</td>
<td>[38]</td>
</tr>
<tr>
<td>rs702634</td>
<td>ARL15</td>
<td>5</td>
<td>1.06</td>
<td>AVG</td>
<td>0.71</td>
<td>[35]</td>
</tr>
<tr>
<td>rs720877</td>
<td>BCART</td>
<td>16</td>
<td>1.12</td>
<td>T/G</td>
<td>0.89</td>
<td>[46]</td>
</tr>
<tr>
<td>rs243021</td>
<td>BCL11A</td>
<td>2</td>
<td>1.08</td>
<td>AVG</td>
<td>0.48</td>
<td>[13]</td>
</tr>
<tr>
<td>rs12454712</td>
<td>BCL2</td>
<td>18</td>
<td>1.09</td>
<td>T/C</td>
<td>0.63</td>
<td>[68]</td>
</tr>
<tr>
<td>rs7172432</td>
<td>C2CD4A/B</td>
<td>15</td>
<td>1.14</td>
<td>AVG</td>
<td>0.58</td>
<td>[44]</td>
</tr>
<tr>
<td>rs1048886</td>
<td>C6ORFS7</td>
<td>6</td>
<td>1.54</td>
<td>G/A</td>
<td>0.15</td>
<td>[43]</td>
</tr>
<tr>
<td>rs2975760</td>
<td>CAPN10</td>
<td>2</td>
<td>1.17</td>
<td>C/T</td>
<td>0.15</td>
<td>[14,16]</td>
</tr>
<tr>
<td>rs3792267</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs7580963</td>
<td>CDKN2A/G8A</td>
<td>9</td>
<td>1.20</td>
<td>T/C</td>
<td>0.80</td>
<td>[13,26,27,39,41,63,89,90]</td>
</tr>
<tr>
<td>rs10965250</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2383208</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs7018475</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs564398</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1552224</td>
<td>CENTD2</td>
<td>11</td>
<td>1.14</td>
<td>A/C</td>
<td>0.87</td>
<td>[13]</td>
</tr>
<tr>
<td>rs13292136</td>
<td>CHCHD9</td>
<td>9</td>
<td>1.11</td>
<td>C/T</td>
<td>0.93</td>
<td>[13]</td>
</tr>
<tr>
<td>rs10401969</td>
<td>CLP2</td>
<td>19</td>
<td>1.13</td>
<td>C/T</td>
<td>0.10</td>
<td>[46,88]</td>
</tr>
<tr>
<td>rs3794991</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs5945226</td>
<td>DUSP9</td>
<td>X</td>
<td>1.18</td>
<td>A/G</td>
<td>0.78</td>
<td>[13,39]</td>
</tr>
<tr>
<td>rs17106184</td>
<td>RAF1</td>
<td>1</td>
<td>1.10</td>
<td>G/A</td>
<td>0.91</td>
<td>[35]</td>
</tr>
<tr>
<td>rs12010175</td>
<td>FAM58A</td>
<td>X</td>
<td>1.21</td>
<td>G/A</td>
<td>0.94</td>
<td>[39]</td>
</tr>
<tr>
<td>rs6017317</td>
<td>FITM2</td>
<td>20</td>
<td>1.09</td>
<td>T/C</td>
<td>0.18</td>
<td>[36]</td>
</tr>
<tr>
<td>rs8050136</td>
<td>FTO</td>
<td>16</td>
<td>1.17</td>
<td>A/C</td>
<td>0.46</td>
<td>[13,27,29,31,46,63,90,92]</td>
</tr>
<tr>
<td>rs9939609</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs11642841</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs9936385</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs6467136</td>
<td>GCC1</td>
<td>7</td>
<td>1.11</td>
<td>G/A</td>
<td>0.50</td>
<td>[36]</td>
</tr>
<tr>
<td>rs7041847</td>
<td>GLIS3</td>
<td>9</td>
<td>1.10</td>
<td>AVG</td>
<td>0.55</td>
<td>[36,39]</td>
</tr>
<tr>
<td>rs1084916</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs11877592</td>
<td>GRSN</td>
<td>9</td>
<td>1.15</td>
<td>AVG</td>
<td>0.54</td>
<td>[94]</td>
</tr>
<tr>
<td>rs3923113</td>
<td>GRB14</td>
<td>2</td>
<td>1.09</td>
<td>A/C</td>
<td>0.59</td>
<td>[38,46]</td>
</tr>
<tr>
<td>rs13389219</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs10866471</td>
<td>GRK5</td>
<td>10</td>
<td>1.12</td>
<td>C/T</td>
<td>0.52</td>
<td>[39]</td>
</tr>
<tr>
<td>rs1111875</td>
<td>HHEX</td>
<td>10</td>
<td>1.13</td>
<td>C/T</td>
<td>0.58</td>
<td>[13,25,26,89,90]</td>
</tr>
<tr>
<td>rs5015480</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs7178572</td>
<td>HKG20A</td>
<td>15</td>
<td>1.09</td>
<td>C/T</td>
<td>0.58</td>
<td>[38,46,90]</td>
</tr>
<tr>
<td>rs717055</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1531343</td>
<td>HMG21</td>
<td>12</td>
<td>1.10</td>
<td>C/T</td>
<td>0.68</td>
<td>[36,39]</td>
</tr>
<tr>
<td>rs7957197</td>
<td>HNF1A</td>
<td>12</td>
<td>1.07</td>
<td>C/T</td>
<td>0.85</td>
<td>[13,41]</td>
</tr>
<tr>
<td>rs7305618</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs4430796</td>
<td>HNF18</td>
<td>17</td>
<td>1.19</td>
<td>G/A</td>
<td>0.51</td>
<td>[39,95,96]</td>
</tr>
<tr>
<td>rs4812829</td>
<td>HNF4A</td>
<td>20</td>
<td>1.09</td>
<td>A/G</td>
<td>0.16</td>
<td>[38]</td>
</tr>
<tr>
<td>rs2334499</td>
<td>HCCA2</td>
<td>11</td>
<td>1.35</td>
<td>T/C</td>
<td>0.45</td>
<td>[34]</td>
</tr>
<tr>
<td>rs4402960</td>
<td>IGFBP2</td>
<td>3</td>
<td>1.14</td>
<td>T/G</td>
<td>0.30</td>
<td>[13,26,27,32]</td>
</tr>
<tr>
<td>rs1470579</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2943641</td>
<td>IRS1</td>
<td>2</td>
<td>1.19</td>
<td>C/T</td>
<td>0.63</td>
<td>[13,51,97]</td>
</tr>
<tr>
<td>rs7578326</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs864745</td>
<td>IAZF1</td>
<td>7</td>
<td>1.10</td>
<td>T/C</td>
<td>0.49</td>
<td>[13,63]</td>
</tr>
<tr>
<td>rs849134</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(continued overleaf)
<table>
<thead>
<tr>
<th>Variants</th>
<th>Candidate gene</th>
<th>Chr</th>
<th>OR</th>
<th>R/NR</th>
<th>RAF</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs5219</td>
<td>KCN11</td>
<td>11</td>
<td>1.14</td>
<td>T/C</td>
<td>0.47</td>
<td>[27,29,92,98]</td>
</tr>
<tr>
<td>rs5215</td>
<td></td>
<td></td>
<td>1.14</td>
<td>C/T</td>
<td>0.40</td>
<td></td>
</tr>
<tr>
<td>rs2237892</td>
<td>KCNQ1</td>
<td>11</td>
<td>1.25–1.32</td>
<td>C/T</td>
<td>0.92 (0.61, 0.67)</td>
<td>[13,32,33,89,99]</td>
</tr>
<tr>
<td>rs231362</td>
<td></td>
<td></td>
<td>1.08</td>
<td>G/A</td>
<td>0.52</td>
<td></td>
</tr>
<tr>
<td>rs2237895</td>
<td></td>
<td></td>
<td>1.29</td>
<td>C/T</td>
<td>NA (0.33)</td>
<td></td>
</tr>
<tr>
<td>rs1535000</td>
<td>KCNK16</td>
<td>6</td>
<td>1.08</td>
<td>T/G</td>
<td>0.47 (0.42)</td>
<td>[36]</td>
</tr>
<tr>
<td>rs972283</td>
<td>KLF14</td>
<td>7</td>
<td>1.07</td>
<td>G/A</td>
<td>0.55</td>
<td>[13]</td>
</tr>
<tr>
<td>rs10842994</td>
<td>KLHC5</td>
<td>12</td>
<td>1.10</td>
<td>C/T</td>
<td>0.80</td>
<td>[46]</td>
</tr>
<tr>
<td>rs8090011</td>
<td>LAMA1</td>
<td>18</td>
<td>1.13</td>
<td>G/C</td>
<td>0.32</td>
<td>[90]</td>
</tr>
<tr>
<td>rs6808574</td>
<td>LPP</td>
<td>3</td>
<td>1.07</td>
<td>C/T</td>
<td>0.60</td>
<td>[35]</td>
</tr>
<tr>
<td>rs6815464</td>
<td>MAEA</td>
<td>4</td>
<td>1.13</td>
<td>C/G</td>
<td>NA (0.58)</td>
<td>[36,37]</td>
</tr>
<tr>
<td>rs7656416</td>
<td></td>
<td></td>
<td>1.15</td>
<td>C/T</td>
<td>NA (0.68)</td>
<td></td>
</tr>
<tr>
<td>rs12970134</td>
<td>MC4R</td>
<td>18</td>
<td>1.08</td>
<td>A/G</td>
<td>0.28</td>
<td>[46]</td>
</tr>
<tr>
<td>rs791595</td>
<td>MIR129-LEP</td>
<td>7</td>
<td>1.17</td>
<td>A/G</td>
<td>0.08</td>
<td>[94]</td>
</tr>
<tr>
<td>rs4275659</td>
<td>MPO/HSP9</td>
<td>12</td>
<td>1.06</td>
<td>C/T</td>
<td>0.67</td>
<td>[35]</td>
</tr>
<tr>
<td>rs1387153</td>
<td>MNTR1B</td>
<td>11</td>
<td>1.09</td>
<td>T/C</td>
<td>0.27</td>
<td>[13,50,100]</td>
</tr>
<tr>
<td>rs10830963</td>
<td></td>
<td></td>
<td>1.03</td>
<td>G/C</td>
<td>0.30</td>
<td></td>
</tr>
<tr>
<td>rs10923931</td>
<td>NOTCH2/ADAM30</td>
<td>1</td>
<td>1.13</td>
<td>T/G</td>
<td>0.09</td>
<td>[63]</td>
</tr>
<tr>
<td>rs35658966</td>
<td>PAM</td>
<td>5</td>
<td>1.23</td>
<td>G/A</td>
<td>0.02</td>
<td>[47]</td>
</tr>
<tr>
<td>rs10229583</td>
<td>PAX4</td>
<td>7</td>
<td>1.14</td>
<td>G/</td>
<td>0.83</td>
<td>[45]</td>
</tr>
<tr>
<td>rs3786897</td>
<td>PED</td>
<td>19</td>
<td>1.10</td>
<td>A/G</td>
<td>0.61 (0.56)</td>
<td>[36]</td>
</tr>
<tr>
<td>rs3130501</td>
<td>POUSF1-TCF19</td>
<td>6</td>
<td>1.07</td>
<td>G/A</td>
<td>0.74</td>
<td>[35]</td>
</tr>
<tr>
<td>rs1801282</td>
<td>PARG</td>
<td>3</td>
<td>1.14</td>
<td>G/C</td>
<td>0.90</td>
<td>[13,23,26,29]</td>
</tr>
<tr>
<td>rs13081389</td>
<td></td>
<td></td>
<td>1.24</td>
<td>A/G</td>
<td>0.95</td>
<td></td>
</tr>
<tr>
<td>rs8042680</td>
<td>PRC1</td>
<td>15</td>
<td>1.07</td>
<td>A/C</td>
<td>0.26</td>
<td>[13]</td>
</tr>
<tr>
<td>rs831571</td>
<td>PSMD6</td>
<td>3</td>
<td>1.09</td>
<td>C/T</td>
<td>0.77 (0.61)</td>
<td>[36]</td>
</tr>
<tr>
<td>rs17584499</td>
<td>PTPRD</td>
<td>9</td>
<td>1.57</td>
<td>T/C</td>
<td>0.23 (0.07)</td>
<td>[99]</td>
</tr>
<tr>
<td>rs7403531</td>
<td>RASGRP1</td>
<td>15</td>
<td>1.1</td>
<td>T/C</td>
<td>0.28 (0.34)</td>
<td>[39]</td>
</tr>
<tr>
<td>rs7560163</td>
<td>RND3</td>
<td>2</td>
<td>1.33</td>
<td>G/C</td>
<td>1.00 (0.87)</td>
<td>[40]</td>
</tr>
<tr>
<td>rs7593730</td>
<td>RBMS1</td>
<td>2</td>
<td>1.11</td>
<td>T/C</td>
<td>0.83</td>
<td>[101]</td>
</tr>
<tr>
<td>rs9552911</td>
<td>SGCG</td>
<td>13</td>
<td>1.5</td>
<td>G/A</td>
<td>0.92</td>
<td>[102]</td>
</tr>
<tr>
<td>rs13342232</td>
<td>SLC16A11-SLC16A13</td>
<td>17</td>
<td>1.29</td>
<td>G/A</td>
<td>0.02 (0.28)</td>
<td>[94]</td>
</tr>
<tr>
<td>rs312457</td>
<td></td>
<td></td>
<td>1.20</td>
<td>G/A</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>rs13266634</td>
<td>SLC30A8</td>
<td>8</td>
<td>1.12</td>
<td>C/T</td>
<td>0.76</td>
<td>[13,25,26,29,89,90,100]</td>
</tr>
<tr>
<td>rs11558471</td>
<td></td>
<td></td>
<td>1.15</td>
<td>A/G</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td>rs3802177</td>
<td></td>
<td></td>
<td>1.26–1.58 (obese/lean)</td>
<td>G/A</td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td>rs1359790</td>
<td>Seyer2</td>
<td>13</td>
<td>1.15</td>
<td>G/A</td>
<td>0.73 (0.71)</td>
<td>[42]</td>
</tr>
<tr>
<td>rs391300</td>
<td>SRR</td>
<td>17</td>
<td>1.28</td>
<td>G/A</td>
<td>0.63 (0.63)</td>
<td>[99]</td>
</tr>
<tr>
<td>rs4523957</td>
<td></td>
<td></td>
<td>1.27</td>
<td>T/A</td>
<td>0.62 (0.65)</td>
<td>[35]</td>
</tr>
<tr>
<td>rs9505118</td>
<td>SSR1-RREB1</td>
<td>6</td>
<td>1.06</td>
<td>A/G</td>
<td>0.61</td>
<td>[35]</td>
</tr>
<tr>
<td>rs16861329</td>
<td>ST6GAL1</td>
<td>3</td>
<td>1.09</td>
<td>G/A</td>
<td>0.88 (0.75)</td>
<td>[38]</td>
</tr>
<tr>
<td>rs7903146</td>
<td>TCF7L2</td>
<td>10</td>
<td>1.4</td>
<td>T/C</td>
<td>0.28</td>
<td>[13,19,25–29,63,89,90,103]</td>
</tr>
<tr>
<td>rs4505655</td>
<td></td>
<td></td>
<td>1.34</td>
<td>T/A</td>
<td>0.30</td>
<td>92,93,103]</td>
</tr>
<tr>
<td>rs7901695</td>
<td></td>
<td></td>
<td>1.37</td>
<td>C/T</td>
<td>0.28 (0.04)</td>
<td></td>
</tr>
<tr>
<td>rs7578597</td>
<td>THADA</td>
<td>2</td>
<td>1.15</td>
<td>T/C</td>
<td>0.88</td>
<td>[63]</td>
</tr>
<tr>
<td>rs2796441</td>
<td>TLE1</td>
<td>9</td>
<td>1.07</td>
<td>G/A</td>
<td>0.61</td>
<td>[46]</td>
</tr>
<tr>
<td>rs6813195</td>
<td>TEM154</td>
<td>4</td>
<td>1.08</td>
<td>C/T</td>
<td>0.72</td>
<td>[35]</td>
</tr>
<tr>
<td>rs998451</td>
<td>TEM163</td>
<td>2</td>
<td>1.56</td>
<td>G/A</td>
<td>0.86</td>
<td>[104]</td>
</tr>
<tr>
<td>rs896858</td>
<td>TPS3INP1</td>
<td>8</td>
<td>1.06</td>
<td>T/C</td>
<td>0.44</td>
<td>[13]</td>
</tr>
<tr>
<td>rs7961581</td>
<td>TSPAN8-LGR5</td>
<td>12</td>
<td>1.09</td>
<td>C/T</td>
<td>0.25</td>
<td>[63]</td>
</tr>
<tr>
<td>rs6705689</td>
<td>UBE2E2</td>
<td>3</td>
<td>1.21</td>
<td>G/A</td>
<td>0.91 (0.83)</td>
<td>[44]</td>
</tr>
<tr>
<td>rs1802295</td>
<td>VPS26A</td>
<td>10</td>
<td>1.08</td>
<td>A/G</td>
<td>0.35 (0.26)</td>
<td>[38]</td>
</tr>
<tr>
<td>rs10010131</td>
<td>WFS1</td>
<td>4</td>
<td>1.11</td>
<td>G/A</td>
<td>0.68</td>
<td>[13,105,106]</td>
</tr>
<tr>
<td>rs6446482</td>
<td></td>
<td></td>
<td>1.11</td>
<td>G/C</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td>rs1801214</td>
<td></td>
<td></td>
<td>1.13</td>
<td>T/C</td>
<td>0.74</td>
<td></td>
</tr>
<tr>
<td>rs4457053</td>
<td>ZBED3</td>
<td>5</td>
<td>1.08</td>
<td>G/A</td>
<td>0.26</td>
<td>[13]</td>
</tr>
<tr>
<td>rs9470794</td>
<td>ZFAND3</td>
<td>6</td>
<td>1.12</td>
<td>C/T</td>
<td>0.12 (0.27)</td>
<td>[36]</td>
</tr>
<tr>
<td>rs11634397</td>
<td>ZFAND6</td>
<td>15</td>
<td>1.06</td>
<td>G/A</td>
<td>0.64</td>
<td>[13]</td>
</tr>
<tr>
<td>rs12571751</td>
<td>ZNF1</td>
<td>10</td>
<td>1.08</td>
<td>A/G</td>
<td>0.53</td>
<td>[46]</td>
</tr>
</tbody>
</table>

R/NR, risk/nonrisk allele; RAF, risk allele frequency in HapMap CEU individuals with risk allele frequency in reference study for non-Europeans.

*In the discovery article RAF for the European population is stated as 0.31.*
names of candidate genes located close to them, only a few are close to strong biologic candidates, including the melatonin receptor (MTNR1B) and the insulin receptor substrate-1 (IRS1). For others, like TCF7L2 and GIPR, the evidence is quite strong that an intronic SNP is the causal SNP. Melatonin receptor 1B (MTNR1B) has been found to be associated with both fasting glucose and T2DM risk [48–50]. Melatonin works as a chronobiotic factor, adjusting the timing of the biologic clock. Its receptors are present in the pancreas and melatonin is proposed to contribute to the nocturnal lowering of insulin in humans. The MTNR1B risk genotype is associated with impaired early insulin release to both oral and intravenous glucose and insulin secretion deteriorates over time in the risk allele carriers [48]. The proposed mechanism by which MTNR1B polymorphism could predispose to T2DM involves altered expression of MTNR1B in pancreatic β-cells leading to decreased cAMP/cGMP concentrations via G proteins and, thereby, impaired insulin secretion.

The insulin receptor substrate 1 (IRS1) gene encodes a protein that mediates insulin’s control of various cellular processes by transmitting signals from the insulin receptor to intracellular signaling pathways. The C allele of rs2943641 has been shown to be associated with insulin resistance and increased risk of diabetes. The genetic variant causes reduced basal levels of IRS1 protein and decreased insulin induction of IRS1-associated phosphatidylinositol-3-hydroxykinase activity in human skeletal muscle biopsies [51].

TCF7L2 is a transcription factor playing an important role in the Wnt signaling pathway. The risk allele is associated with decreased insulinogenic index and lower disposition index, suggesting a reduced capacity for insulin secretion in relation to insulin sensitivity. Since it was identified as a diabetes gene it has been shown to be important for several vital functions in the pancreatic islet, including pancreas development, determination of β-cell mass, and maintenance of the secretory function of mature β-cells.

The incretin hormone GIP (glucose-dependent insulino tropic polypeptide) promotes pancreatic β-cell function by potentiating insulin secretion and β-cell proliferation. The GIP receptor (GIPR) locus showed association to postprandial insulin levels in a meta-analysis performed by the MAGIC consortium but was surprisingly not associated with risk of diabetes in the DIAGRAM+ study [13,52]. The reason seems to be that the same variant results in decreased BMI, which neutralizes the effect of the SNP on risk of T2DM. GIP influences expression of the inflammatory cytokine OPN in islets and fat, which in turn, has protective effects on β-cell proliferation and potentially apoptosis, but detrimental effects on insulin sensitivity [53,54].

Many of the other recently identified loci can be subgrouped based on their association with other phenotypes with a key role in T2DM etiology. Exploration of the effects of T2DM-associated variants on glucose and insulin traits in nondiabetic populations has shown that most of the known loci act through an effect on insulin secretion rather than insulin resistance (Table 26.1) [13,24,55,56].

Fasting glucose-raising alleles of the MADD, GIPR, GCK, FADS, DGKB, PROXI, TCF7L2, SLC30A8, HHEX/IDE, CDKAL1, CDKN2A/2B, and C2CD4B loci have all been associated with either abnormal insulin processing or secretion whereas GCKR and IGF1 are associated with OGTT-based disposition indices and β-cell function [55,57]. The DIAGRAM+ consortium observed that three loci (TCF7L2, ARAP1, and CDKAL1) were associated with reduced fasting insulin also suggestive of β-cell dysfunction, whereas the T2DM risk alleles at PPARG, FTO, IRS1, GCKR, and KLF14 were associated with higher fasting insulin, indicating a primary effect on insulin action [13,57].

The missing heritability

In spite of the large number of risk variants identified, it is estimated that they explain less than 15% of the heritability of T2DM. The unexplained heritability is an intensely discussed topic in complex genetics, some claiming it as a failure of GWA studies. There are many possible explanations for the missing heritability, including assumptions made about the genetic architecture of the disease and the definitions of heritability. The estimations of heritability explained assumes that only additive effects determine disease risk and that the risk follows the liability threshold model, that is, the genetic and environmental effects sum up to form a normal distribution of liability and that disease arises in individuals surpassing a certain threshold in the distribution [58]. If these assumptions are not true, the estimate of heritability explained will not be correct. However, there are also many other potential explanations for the missing heritability: yet unmapped common variants, distorted parent-of-origin transmission of risk alleles, rare variants, structural polymorphisms (e.g. copy number variations), gene–gene and/or gene–environment interactions (in which epigenetic effects may be important).

The genetic architecture of T2DM

There is an ongoing debate about the genetic architecture underlying common complex diseases. Some argue that disease risk is determined by a large number of very small additive effects from relatively common variants and that disease represents the extremes of a normal distribution [59] while others argue that complex diseases are rather collections of phenocopies caused by rare, often recent, mutations [60]. One argument against common variants is that they would have been removed from the population by natural selection [8]; however, this is not a valid argument for a disease like T2DM where the penetrance of the genetic effect depends strongly on interactions with the environment, especially since this environment has changed in recent years and the genetic risk variants could have been neutral or even beneficial before the
introduction of the Westernized lifestyle. Applying an approach that considers all SNPs in a GWAS could in fact explain a much larger proportion of the heritability (>50%) supporting the existence of numerous yet unidentified loci with smaller effects [13,61]. However, one should keep in mind that heritability can only be estimated from the most recent generations for which information on affection status is available, whereas most of the variants studied thus far are ancestral variants hundreds of generations old. We do not know whether these ancestral variants (which have modest effects and have escaped purifying selection) can really explain the diabetes epidemic we see in the most recent generations or whether this can be ascribed to rare variants with stronger effects. The truth may thus rather be a combination of the suggested models, at least for T2DM.

The genetic architecture could also vary within the T2DM patient group since the diagnosis may include cases of disease caused by rare, or even unique, variants with high penetrance, in parallel or combination with cases pushed over the disease threshold by their load of common risk alleles. Allelic heterogeneity can also be expected at any single disease locus, that is, there may be multiple, different susceptibility mutations at the locus conferring risk in different individuals and both common and rare variants could contribute to disease susceptibility similar to what has been found at the HNF1A and HNF4A MODY loci. Structural polymorphisms and microRNAs add a further layer of complexity and have not yet been exhaustively studied.

The rapid development of next generation sequencing tools has markedly facilitated discovery of rare variants. From being both extremely expensive and effort consuming, sequencing of the whole genome with appropriate coverage (approx. 30 times) can now be performed for <$1000 in a few days. Whole-exome sequencing can be performed for less than $400 per sample and these figures are likely to decrease over the coming years. Applying these methods in families and large population studies will hopefully answer the questions about the role of rare variants in complex diseases.

Most studies to date have been limited to SNP leaving structural polymorphisms relatively unexplored. However, since common structural variants are likely to be tagged by surrounding SNPs they are unlikely to explain a large proportion of missing heritability. A recent study [62] identified a common copy number variation (CNV), CNVR5583.1 (TSPAN8), as associated with T2DM. This association could be convincingly replicated by previously typed SNPs that tag the CNV [63].

Noncoding RNAs have recently emerged as important regulators of gene expression and function. MicroRNAs (miRNAs) naturally regulate programs of gene expression. Altered miRNA function has been shown to contribute to human disease, and manipulation of specific miRNAs is now being explored as a novel therapeutic modality [64]. The efficiency of miRNAs binding to target transcripts depends on both the sequence and the intramolecular structure of the transcript. SNPs can contribute to alterations in the structure of regions flanking them, thereby influencing the accessibility for miRNA binding. Several studies have implicated miRNAs in diabetes and inflammation and common SNPs change the target sequence of miRNAs in several T2DM susceptibility loci [65,66]. Other forms of noncoding RNAs, such as piRNAs (PIWI-interacting RNAs), snoRNAs (small nucleolar RNAs), lncRNAs (long intergenic noncoding RNAs), and IncRNAs (long noncoding RNAs), may also contribute to the development of diabetes. For example, the CDKN2A/B region on chromosome 9 is associated with T2DM, as well as cardiovascular disease and a number of other disorders. This region harbors an IncRNA, ANRIL (nonprotein coding CDKN2B-AS1 CDKN2B antisense RNA 1), which can potentially modify and explain some of these associations [67].

Gene–gene and gene–environment interactions

Gene–gene interactions, or epistasis, have been suggested as a possible explanation for difficulties in replicating genetic association in complex diseases [68]. The standard statistical methods used in association studies are usually limited to analysis of single marker effects and thereby do not account for interactions between markers. Previous attempts to study epistasis in complex diseases have focused on interactions between candidate regions [69,70]. However, the recent abundance of GWAS data has made a comprehensive search across the genome more feasible. Some studies have attempted to account for epistasis in GWAS using a two-step approach in which significant SNPs are tested against each other or against all other SNPs in the study with variable results [71,72]. The main problem when studying epistasis is power, since interaction between loci with modest effects is difficult to detect without extremely large sample sizes. However, a study using simulated data has shown that power could actually increase when testing all pair-wise combinations of SNPs in GWAS settings despite the penalty for multiple testing; especially when no marginal effects were present [73]. Thorough studies in diabetes addressing epistasis using this approach are missing.

Further, a recent paper by Eric Lander and coworkers provided compelling evidence that gene–gene interaction can also contribute to missing heritability by causing “phantom heritability” that inflates the estimated narrow sense heritability of the trait [74].

Gene–environment interactions are equally difficult to study but are likely to play an important role in T2DM development. The epidemic of T2DM only dates back 50 years, and it is quite obvious that during this period only the environment, not the genes have changed. However, the genetic architecture determines our response to the environment. Genetic variants could affect specific metabolic processes to make an individual more susceptible to the harmful effects of a poor diet but also personality traits that make an individual more or less likely to overconsume and live a sedentary lifestyle. It will however be a formidable task to identify the environmental triggers for most of the genetic variants increasing susceptibility to diabetes as
this will require very large studies with precise information on diet, exercise, energy expenditure, and so on.

**Epigenetics**

The environment can also influence the expression of the genome, and ultimately the phenotype, via the epigenome. Even though the DNA sequence is not changed, the phenotype is altered by epigenetic modifications of gene expression by mechanisms including methylation of DNA, posttranslational modification of histones, or activation of microRNAs. These modifications have the potential to be stable and heritable across cell divisions [75,76]. Changes to the phenotype can be at the level of the cell, tissue, or whole organism.

It is tempting to speculate that environmental factors such as diet and exercise can change the level of DNA methylation and thereby cause changes in gene expression, but evidence that DNA methylation contributes to the increase in T2DM is still lacking. Epigenetic mechanisms may, however, play a role in progression of the disease by inducing glucotoxicity in islets and predispose to diabetic complications. Elevated glucose is a prerequisite to this condition and it is well established that cells can memorize changes in glucose concentrations. For example, two large studies, the UKPDS and DCCT studies, showed that an initial good metabolic control was associated with reduced frequency of diabetic complications decades later. The advanced “metabolic memory” hypothesis suggests that this is because glucose can induce histone modifications in endothelial cells that can be remembered long after [77].

As previously mentioned, the risk of T2DM in offspring is greater if the mother has T2DM compared to if the father is affected. The reasons for this parent-of-origin effect are unknown but one potential explanation could be distorted parent-of-origin transmission of risk alleles which is often associated with DNA methylation and imprinting. Variants in both KCNQ1 and KLF14 show stronger effects on T2DM when the risk allele is transmitted from the mother than from the father [34,78]; in some instances the paternal allele can even be protective making it almost impossible to detect association in a traditional case-control study.

Epigenetic mechanisms can also act in utero. If this intrauterine programming results in a reduced β-cell mass, it could predispose to diabetes later in life when the insulin requirements increase as a consequence of obesity and insulin resistance.

**Systems biology**

GWAS on their own provide limited insights into the molecular mechanisms driving disease. To reach an understanding of disease pathogenesis, it is important to analyze the GWAS data in the context of complementary types of follow-up analyses such as related protein module analysis, expression profiling under conditions relevant for the disease, and analysis of genotype-phenotype associations [79]. For example, by combining GWAS information with metabolomics it has been possible to identify strong associations between SNPs and metabolic reactions that otherwise would have been missed [80]. Network or pathway-based approaches, including enrichment in pre-defined pathways by, for example, KEGG [81] (www.genome.jp) and Gene Ontology (GO) (www.geneontology.org), have also been used to identify disease genes for various diseases [79,82–85]. Thus, an integrative approach of several data types is likely to discover disease genes that would not be identified by the use of classical GWAS approaches. This was also illustrated in a recent study of human islets identifying novel candidate genes for T2DM based upon expression differences and co-expression with known T2DM genes as well as protein–protein interaction analyses [86]. Integration of GWAS data with such data could, thus, facilitate a systems-based understanding of the pathogenic mechanisms.

**Conclusions**

The technical revolution in the field of genetics has allowed identification of numerous genetic variants that associate with T2DM. Yet, the dissection of the genetics of T2DM is still in its infancy, so far only explaining a small proportion of the total heritability of diabetes. In spite of this, it has already greatly contributed to our understanding of disease mechanisms by identifying pathways that could not be linked to diabetes by existing hypothetical models, even though many genetic findings are very recent and have yet to make their contribution to our knowledge about diabetes pathogenesis. Diabetes is probably a much more diverse disease than the current subdivision into T1DM and T2DM implies and a more precise subdivision into subgroups may both facilitate the investigation of T2DM genetics and pave the way for more individualized treatment. A holistic systems biology approach will also be required to obtain a complete picture of how genetic variation leads to diabetes. The rapid technology development during the past years holds promises that this will be possible in a not too distant future.

**References**

5. Poulsen P, Kyvik KO, Vaag A, Beck-Nielsen H: Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal


76 Brasacchio D, Okabe J, Tikellis C, et al.: Hyperglycemia induces a dynamic cooperativity of histone methylase and demethylase


93 The Wellcome Trust Case Control Consortium: Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007;447(7145):661–678.


CHAPTER 27
Glucose toxicity

Hannele Yki-Järvinen¹ and Donald A. McClain²
¹Department of Medicine, University of Helsinki, Minerva Foundation Institute for Medical Research, Helsinki, Finland
²Department of Medicine, University of Utah and Veterans Affairs Medical Center, Salt Lake City, UT, USA

Key points
• Acute hyperglycemia increases glucose utilization (mass-action effect of glucose).
• Chronic hyperglycemia impairs glucose utilization (“glucose toxicity”).
• Acute hyperglycemia stimulates insulin secretion.
• Chronic hyperglycemia impairs insulin secretion (“glucose toxicity”).
• The rate of glucose utilization is decreased in type 1 and type 2 diabetic patients when studied under normoglycemic conditions but normal when measured under hyperglycemic conditions.
• Any intervention, which improves glycemic control, improves insulin sensitivity and insulin secretion.
• The hexosamine pathway has been convincingly linked to glucose-induced insulin resistance and β-cell failure.
• The hexosamine pathway metabolizes a small fraction of glucose to UDP-N-acetylglucosamine, which responds to increased cellular glucose flux in insulin-dependent tissues by downregulating insulin-stimulated glucose uptake.

Introduction
Several studies, culminating in the large and definitive Diabetes Control and Complications Trial and the United Kingdom Prospective Diabetes Study [1,2] have established that hyperglycemia is the major risk factor for microvascular diabetic complications. Chronic hyperglycemia also seems to be a significant and independent, albeit weaker risk factor for macrovascular disease [2]. Other adverse consequences of hyperglycemia include an increased susceptibility to infections [3]. These complications affect tissues wherein glucose utilization is predominantly non-insulin-dependent. In addition, high concentrations or flux rates of glucose can exert adverse metabolic consequences on insulin-dependent tissues that regulate glucose disposal. High concentrations of glucose can by themselves cause two of the principal hallmarks of type 2 diabetes, namely deficiencies of insulin secretion and insulin action, both in animals [4,5] and in humans [6].

The term “glucose toxicity” was first applied to these phenomena by Rossetti and DeFronzo in 1990 [7], although the term had been used earlier to describe other adverse effects of hyperglycemia. In this review, we will restrict the use of the term to the ability of excess glucose to alter normal glucose homeostasis itself, and further restrict the discussion to those effects that are mediated by metabolism of glucose through its normal pathways. Thus, we will exclude effects of hyperglycemia that are, for example, mediated by nonenzymatic glycation or hyperosmolarity, even though these are of unquestioned importance in diabetes. We also acknowledge that the use of the term “toxicity” may not be the most appropriate for at least two reasons: Firstly and most obviously, glucose does not fit most definitions of “toxin,” and secondly, some of the “toxic” manifestations of glucose such as insulin resistance can also be seen as adaptive mechanisms that actually protect the organism from other, potentially even worse consequences of high nutrient fluxes. We also emphasize that glucose toxicity is but one of many pathogenetic mechanisms operative in type 2 diabetes. There are other routes to insulin resistance and β-cell failure such as inflammation, lipotoxicity, and others that operate in concert with glucose toxicity to create and maintain the full diabetic phenotype.

In this review, after a brief summary of the pathophysiology of hyperglycemia, we will review the evidence that hyperglycemia can cause insulin resistance and impaired insulin secretion. We will then consider the underlying molecular mechanisms for, and finally the clinical relevance of, glucose toxicity.
Hyperglycemia and glucose utilization

The mass-action effect of glucose
In virtually all tissues except the brain, glucose, at a fixed insulin concentration, promotes its own utilization in a concentration-dependent manner (Figure 27.1) [8]. In insulin-sensitive tissues, the glucose-induced increase in glucose utilization is dependent upon the insulin concentration. For example, an increase in the blood glucose concentration from a fasting concentration from 5 to a peak postprandial concentration of 8.9 mmol L\(^{-1}\) (from 90 to 160 mg dL\(^{-1}\)) increases the rate of whole-body glucose utilization two- to threefold more in the presence of postprandial (serum insulin 50–160 mU L\(^{-1}\)) than fasting (insulin 10–20 mU L\(^{-1}\)) serum insulin concentrations [8]. This mass-action effect of glucose is quite significant if one compares it to the rate of whole-body glucose utilization which averages \(\sim 2\) mg kg\(^{-1}\) min\(^{-1}\) after an overnight fast, and increases two- to threefold to \(\sim 5–7\) mg kg\(^{-1}\) min\(^{-1}\) after a meal [9]. As discussed later, the mass-action effect of glucose has important implications for the understanding of the mechanisms underlying glucose toxicity.

Consequences of the acute stimulatory effect of hyperglycemia on postprandial glucose disposal in patients with type 1 and 2 diabetes
Whole-body glucose uptake can be accurately quantitated, using the insulin clamp technique during maintenance of similar glucose and insulin concentrations in normal subjects and diabetic patients [10]. Under such conditions, insulin resistance is observed, that is, the rate of glucose uptake is, on average, reduced in both patients with type 1 and 2 diabetes [11]. If, however, the rate of glucose uptake is measured in patients with type 1 [12] or 2 [9] diabetes under conditions simulating the actual postprandial glucose and insulin concentrations found in these patients, the rate of absolute insulin-stimulated glucose uptake is normal. This is explained by the ability of hyperglycemia, via glucose mass-action, to compensate for the reduction in glucose utilization caused by insulin resistance. In the absence of any insulin resistance, the mass-action effect of glucose would lead to overutilization of glucose not only in non-insulin-dependent but also insulin-dependent tissues. In non-insulin-dependent tissues not protected by the blood–brain barrier such as peripheral nerves, the kidney, and retina, the rate of glucose utilization is chronically elevated under hyperglycemic conditions, and is a key pathophysiologic abnormality leading to diabetic microvascular complications [1,2]. In contrast, the normal glucose flux to insulin-sensitive tissues, such as skeletal muscle, spares them from hyperglycemia-induced damage.

Chronic hyperglycemia as a cause of insulin resistance

Insulin resistance in patients with type 1 diabetes—a consequence of glucose toxicity?
Insulin resistance both precedes and predicts type 2 diabetes and therefore is not merely due to hyperglycemia [13]. In the case of type 1 diabetes, it is clear that insulin resistance is an acquired and reversible phenomenon since insulin sensitivity is completely normalized during remission of the disease [14]. In these patients, insulin resistance of glucose utilization is predominantly localized to skeletal muscle [15]. Since the peripheral (although not the portal) insulin concentration is usually similar to that in normal subjects in insulin-treated patients [16], insulin deficiency cannot explain insulin resistance in skeletal muscle.

Induction of in vivo insulin resistance by hyperglycemia
In type 1 diabetic patients lacking endogenous insulin secretion, effects of hyperglycemia per se on insulin action can be examined. In such a study, insulin-stimulated glucose uptake was measured on two occasions, after 24 hours of hyperglycemia induced by an intravenous glucose infusion (mean diurnal blood glucose 16 mmol L\(^{-1}\)) and after 24 hours of normoglycemia (6 mmol L\(^{-1}\)), in the face of maintaining identical diurnal serum insulin profiles by continuous subcutaneous insulin infusion [6,15] (Figure 27.2). After 24 hours of hyperglycemia, the rate of whole-body glucose uptake was consistently lower than after normoglycemia, demonstrating that short-term hyperglycemia can cause insulin resistance in humans [6,15] (Figure 27.2). These observations were subsequently confirmed [17–19].

Despite the strong suggestion that glucose per se was inducing insulin resistance in patients with type 1 diabetes, the complexity of metabolic regulation still left open the possibility that the
Insulin resistance might result from as yet unidentified factors. Animal models, however, gave clearer proof of the primacy of glucose in this phenomenon. In the rat, for example, removal of 90% of the pancreas causes insulin deficiency, hyperglycemia, and a 30% reduction in insulin-stimulated glucose utilization in skeletal muscle [5]. Selective correction of hyperglycemia with phlorizin, which induces glucosuria via inhibition of glucose reabsorption in the proximal tubule, normalizes glycemia without changing serum insulin concentrations (Figure 27.3). These studies together with the human data provided the first evidence of the ability of the blood glucose concentration itself to regulate insulin sensitivity, and the phenomenon began to be referred to as “glucose toxicity” [7,20].

**Hyperglycemia as a mediator of insulin resistance in type 2 diabetes**

In type 2 diabetes, a large number of studies are consistent with the glucose toxicity concept although interpretation of these studies is sometimes not as simple as the studies of type 1 diabetes because of the complex pathophysiology of type 2 diabetes. For example, it is known that hyperinsulinemia can cause insulin resistance [21,22], and type 2 diabetic patients early in the course of the disease are often hyperinsulinemic. It could be argued, therefore, that any treatment that makes a subject more insulin sensitive will ultimately lower insulin levels, making them more insulin sensitive through that mechanism alone. Likewise, weight loss improves both glycemia and insulin resistance [23] but it is difficult to attribute the causality for the improved insulin sensitivity to glycemia alone in such studies. Nevertheless, the data available are at least consistent with the concept of glucose toxicity. Garvey and coworkers, for example, demonstrated that 3 weeks of intensive insulin therapy significantly improved the maximum glucose disposal rate, endogenous glucose output, and insulin secretory capacity [24], and similar findings were also obtained in other laboratories [25]. The fact that basal insulin levels were unchanged in the weight loss study [23] and increased after intensive insulin therapy [24] suggest that insulinemia per se was not the driving force for the changes in insulin sensitivity. Furthermore, other therapeutic interventions such as sulfonylureas also improve insulin sensitivity, consistent with
the concept that glucose was a principal driver of the insulin resistance prior to therapy.

Since that time, additional support for this concept has been obtained from a wide variety of in vitro and in vivo models, a few examples of which follow. The in vitro studies are particularly compelling because of the ability to isolate glucose as a dependent variable. Insulin-stimulated glucose transport in muscle strips of hyperglycemic type 2 diabetic patients was lower than that of normoglycemic subjects but prolonged exposure to normoglycemia completely reversed that defect [26]. Isolated adipocytes also respond to high concentrations of glucose by desensitizing their glucose transport system [27]. Numerous other seminal studies in animals that support the concept of glucose toxicity as a major determinant of insulin sensitivity have been reviewed in the past and will not all be recapitulated here [7,20].

**Physiologic basis of glucose-induced insulin resistance in vivo**

In insulin-treated patients with type 1 diabetes, in whom glucose toxicity appears to be the major cause of insulin resistance, direct quantitation of tissue glucose uptake during insulin stimulation using positron emission tomography has demonstrated that skeletal muscle is the predominant tissue responsible for the defect in insulin-stimulated glucose utilization [28]. Since a substantial fraction of whole-body glucose oxidation is independent of insulin, rates of whole-body glucose oxidation have to be corrected for the estimated contribution of this component of glucose oxidation [29]. When this is done in patients with type 1 [12] or type 2 [30] diabetes, the fractions of glucose oxidized and disposed of nonoxidatively (the sum of glycogen synthesis and nonoxidative glycolysis) are similar to those in normal subjects. These data suggest that regardless of the primary cellular process causing insulin resistance, the ultimate gate-keeper for cellular glucose uptake is located at the level of glucose transport or phosphorylation. In support of this, muscle glucose-6-phosphate concentrations are similar in insulin-resistant patients with type 1 diabetes under conditions where glucose flux is acutely normalized by hyperglycemia under hyperinsulinemic conditions as under conditions of normoglycemic hyperinsulinemia [12]. Since both GLUT4-mediated glucose transport [31] and glucose phosphorylation by hexokinase II [32] are regulated by insulin, either transport or phosphorylation could be rate-limiting for glucose disposal. Another hypothetical possibility is that glucose delivery could be rate-limiting for glucose disposal. Insulin, at physiologic concentrations, increases blood flow, depending on factors such as limb musculature and physical fitness from −10 to 80% (mean in 75 studies around 20%) [33]. During a similar time period and at similar insulin concentrations, glucose extraction, as determined from the AV-glucose difference across a limb, increased 1000 to 2000% within 30−60 minutes. Although defects in blood flow may be observed at supraphysiologic insulin concentrations in patients with type 1 or type 2 diabetes, studies performed at physiologic insulin concentrations locate impaired insulin action exclusively to glucose extraction [33]. Taken together these data localize hyperglycemia-induced insulin resistance of glucose utilization to early steps in glucose uptake in skeletal muscle. This defect is accompanied by similar relative reductions in glucose oxidation and storage.

**Chronic hyperglycemia as a cause of impaired insulin secretion**

In 1948, Lukens and Dohan administered large doses of glucose to normal cats and induced permanent hyperglycemia, hydropic degeneration of the islet of Langerhans, and ketonuria in 4 out of 35 cats studied [34]. These investigators proposed that hyperglycemia could play a role in the pathogenesis of diabetes. In recent years, several approaches have been used to directly examine the harmful effects of chronic hyperglycemia on insulin secretion. These studies have established that chronic hyperglycemia impairs β-cell responsiveness to glucose although the exact biochemical mechanism(s) mediating this effect remain to be clearly defined.

**Insulin secretion during experimental hyperglycemia**

In normal subjects, β-cell hyperresponsiveness and α-cell hyporesponsiveness is observed after mild hyperglycemia induced by glucose infusions [35]. However, exposure to higher glucose concentrations (12.6 mmol L\(^{-1}\)) for 68 hours has been reported to be associated with significant reduction in insulin secretion in humans [19]. As discussed under section "Clinical significance" later, treatment studies in patients with type 2 diabetes have provided additional support for the idea that chronic hyperglycemia impairs insulin secretion also in humans.

Since the early experiments in cats [34], the ability of chronic hyperglycemia to desensitize the β cell to glucose has been convincingly documented in animal models. In the partially pancreatectomized dog, maintenance of plasma glucose at greater or equal to 14 mmol L\(^{-1}\) for two weeks induces persistent hyperglycemia, loss of glucose induced insulin secretion, ketonuria, and weight loss [36]. Morphometric analysis of the endocrine pancreas revealed a profound reduction in the number and size of islets. These changes were not observed in portions of pancreases removed prior to the glucose infusions or in pancreases of similarly pancreatectomized dogs not subjected to hyperglycemia. Leahy and Weir infused normal rats in vivo with various concentrations of glucose and measured the insulin response thereafter in vitro using the isolated perfused pancreas [37]. Rats infused with 35% glucose showed a severely blunted insulin response to glucose after 48 hours, and in rats infused with 50% glucose for 48 hours, the glucose-induced insulin response was totally lost. Addition of phlorizin to the 50% glucose infusion after 48 hours for an additional 48-hour
Molecular mechanisms of glucose toxicity

Carbons from ingested glucose are incorporated into nearly every class of biologic metabolite and macromolecule, and the metabolism of those molecules generates energy, oxidant byproducts, and inflammatory stress, all of which could contribute to glucose-induced insulin resistance. The complexity of the pathways of insulin signal transduction and its interactions with numerous other regulatory networks suggests that glucose toxicity is unlikely to operate through a single mechanism, and the data are consistent with that presumption. Two leading candidates for the molecular mediators of glucose toxicity are the hexosamine/O-linked glycosylation pathway and redox signaling pathways, and both of these are supported by extensive experimental data that will be summarized in this section. These are unlikely to be the only mechanisms for glucose toxicity, however. Excess glucose is metabolized to lipid, for example, and there is no question that excess lipid fluxes through mechanisms involving activation of protein kinase C, ceramide synthesis, accumulation of incompletely metabolized intermediates, and others can contribute to insulin resistance and β-cell failure in diabetes. For consideration of these pathways and mechanisms of vascular complications of diabetes the reader is referred to the other chapters in this volume and several excellent recent reviews [39,40]. We will concentrate this review on the more direct products of glucose metabolism that contribute specifically to the metabolic abnormalities of diabetes.

The hexosamine/O-linked N-acetyl glucosamine pathway

The hexosamine biosynthetic pathway and insulin resistance

One metabolic fate of glucose is the hexosamine biosynthetic pathway (HBP), and high flux through this pathway, such as occurs in diabetes or overnutrition, has been convincingly linked to insulin resistance and β-cell failure. In this pathway a relatively small fraction of cellular glucose flux—a few percent in most tissues—is converted to UDP-N-acetylglucosamine (UDP-GlcNAc) and other amino sugars (Figure 27.5). The rate-limiting step is catalyzed by the enzyme glutamine:fructose-6-phosphate amidotransferase (GFA) that catalyzes the synthesis of glucosamine-6-P from fructose-6-P and glutamine. The product of the HBP, UDP-GlcNAc is well placed to serve a nutrient-sensing function, as its production has been shown to be responsive not only to cellular glucose flux [27] but also to the levels and availability of nucleotides [41], fatty acids [41], and amino acids [27]. It is a high-energy intermediate and is therefore also reflective of cellular energy status.

The first evidence of the involvement of the HBP in metabolic regulation was its mediation of glucose toxicity in models of insulin resistance. High concentrations of glucose downregulate insulin-stimulated glucose uptake, and Marshall made a breakthrough in understanding this phenomenon when he demonstrated in adipocytes that glucose metabolism to glucosamine-6-P was required for this effect [27]. Since then, numerous studies have shown that increased hexosamine flux can induce insulin resistance in cultured cells and whole animals (reviewed in [42]). Many of these studies employed infusion of glucosamine that enters the HBP directly after uptake by glucose transporters and phosphorylation by hexokinase. Interpretations of these studies were criticized because pharmacologic concentrations of glucosamine can compete for glucose uptake, deplete cellular ATP, and even trigger apoptosis [43]. Furthermore, short-term infusion studies may not reflect the full range of posttranslational and transcriptional changes that can result from chronic activation of the HBP. More definitive evidence for the involvement of the HBP in insulin resistance was obtained using animal models wherein chronic but physiologic changes in hexosamine flux were induced by overexpression of GFA in a tissue-specific manner. The results of these studies showed that
the HBP serves a nutrient-sensing function in multiple tissues and affects metabolism in a wide-ranging manner. Specifically:

- overexpression in muscle plus fat, or fat alone results in insulin resistance, downregulation of glucose transport [44–46], and;
- overexpression in β cells results in hyperinsulinemia [47].

These phenotypes mimicked and therefore validated those that had been seen in the infusion models and with glucosamine treatment of explanted tissues [48–50]. Importantly, the effects of increased HBP flux are nonadditive with hyperglycemia [51] and high-fat feeding [52], in inducing insulin resistance suggesting that both hyperglycemia and overnutrition in general may induce insulin resistance through the HBP.

O-linked glycosylation of nuclear and cytosolic proteins as the mechanism for HBP effects on glucose homeostasis

What is the mediator of the effects of the HBP? UDP-GlcNAc is a substrate for enzymatic glycosylation of proteins, suggesting a mechanism for HBP signaling. Since neither N-linked glycosylation nor O-linked glycosylation of secreted proteoglycans are responsive to changes in glucose fluxes, interest in HBP signaling originally focused on the then recently discovered pathway of O-linked glycosylation wherein single O-GlcNAc residues are enzymatically added to serines and threonines of nuclear and cytosolic proteins (Figure 27.5). Discovered by Hart in 1984, this form of glycosylation is dynamic, is commonly found on proteins involved in signal transduction, and is responsive to changes in glucose flux [53,54]. O-glycosylation is often reciprocal with phosphorylation and in some cases occurs on the same residues that are otherwise phosphorylated (e.g. c-myc [55]), further suggesting a potential role in signaling. Glycosylation is accomplished by a single O-linked GlcNAc transferase (OGT) [56] and deglycosylation is by a single known O-GlcNAcase [57]. The enzymologic and biochemical aspects of the pathway involved in protein O-GlcNAc modification have been reviewed recently and will not be described here in further detail [58]. Importantly and consistent with playing a role in glucose homeostasis, the pathway is highly feedback-regulated by substrates and transcriptional control of OGT and O-GlcNAcase such that in cultured cell lines the minimum levels of glycosylation of at least some proteins

![Figure 27.5 The hexosamine metabolic pathway discovered by Marshall et al. [27]. A small fraction (2–3%) of glucose, which enters insulin-sensitive tissues either using the insulin-independent glucose transporter GLUT1 and glucose phosphorylating enzyme hexokinase I (HK1), or with the help of the insulin-sensitive glucose transporter GLUT4 and hexokinase II (HKII) is metabolized to hexosamine products in the presence of glutamine. GFA is the rate-limiting enzyme of the pathway. The end-product of the pathway is UDP-GlcNAc, which can be covalently linked to serine and threonine residues on specific proteins (O-glycosylation) such as insulin signaling molecules and GLUT4. O-glycosylation seems to decrease insulin action and GLUT4 translocation resulting in insulin resistance. Initial activation of the pathway is thought to occur via glucose mass action, which is at low insulin concentrations after an overnight fast, and is increased under hyperglycemic conditions. Glucosamine can bypass the rate-limiting step of the pathway and can therefore be used as a tool to study consequences of activation of the hexosamine pathway. G–6–P, glucose-6-phosphate; F–6–P, fructose-6-phosphate; TCA, tricarboxylic acid.](http://example.com/figure27.5.png)
happens to occur at the level of normoglycemia in mammals, 5 mM glucose [59]. Consistent with that finding, the pathway regulates insulin signaling even in the normoglycemic range, demonstrating that it is a physiologic and not solely pathophysiologic (“toxic”) regulator [60]. Finally, the pathway’s highly evolutionarily conserved role in nutrient homeostasis and hormone responsiveness in *C. elegans* [47,61], *Drosophila* [62], and plants [63] speaks to its importance in metabolic regulation.

The first direct evidence that the O-glycosylation mediated the observed effects of the HBP on insulin resistance was demonstrated by overexpression of the enzyme responsible for this glycosylation, O-linked GlcNAc transferase (OGT), in skeletal muscle and fat under control of the GLUT4 promoter [64]. This maneuver recapitulated the insulin resistance and hyperleptinemia seen with overexpression of GFA in the same tissues [44]. Other early evidence for the role of OGT came from demonstrations in Hart’s laboratory that enzymatic inhibition of the enzyme O-GlcNAcase in 3T3L1 cells leads to insulin resistance in parallel with increased O-linked glycosylation of proteins [65]. It should also be pointed out, however, that there are models of insulin resistance that do not involve increased levels of O-GlcNAc [66], and conversely, there are circumstances in which increased O-GlcNAc may not lead to insulin resistance [67]. In both of the latter cases, extremely high and unphysiologic levels of protein O-GlcNAc modification were achieved pharmacologically.

Modification of metabolic enzymes and transcription factors by O-GlcNAc has been shown to alter their function in a way that is consistent with the phenotypes observed in animals. Glycogen synthase (GS), the first protein shown to be affected by phosphorylation, is also a target of the HBP/O-GlcNAc pathway [68]. Exposure of cultured adipocytes to high glucose or glucosamine, or streptozotocin-induced diabetes in mice renders the enzyme insulin resistant and less sensitive to its principal positive allosteric regulator, glucose-6-phosphate [69], and this is related not to changes in phosphorylation but rather to levels of O-glycosylation. O-glycosylation of GS therefore mimics phosphorylation and inhibits the enzyme, but in a way that cannot be reversed by insulin action. Thus, O-GlcNAc acts as a dominant negative signal that renders GS insulin resistant and limits glycogen accumulation in situations of chronic hyperinsulinemia and nutrient excess.

Many other components of the insulin signal transduction pathway are modified by O-GlcNAc, most often in an inhibitory fashion. Rats infused with glucosamine, for example, exhibit insulin resistance of skeletal muscle through decreased phosphorylation of insulin receptor substrate-1 (IRS-1) and its association with phosphatidylinositol 3-kinase (PI3K) [56]. IRS-1 [65], the downstream insulin signaling kinase Akt [70], proteins of glucose transporter vesicles [71], the crucial transcriptional metabolic regulator FOXO1 [72], and the nutrient sensor AMP-activated protein kinase (AMPK) [73] are also modified by O-GlcNAc, with proven functional consequences. A recent and particularly exciting insight into the crosstalk between hormone signaling and O-GlcNAc is that OGT contains a domain that recognizes phosphatidylinositol 3,4,5-trisphosphate [74]. This results in the recruitment of OGT to the plasma membrane after activation of PI3K through the insulin/IR/IRS-1 pathway. Thus, even as the insulin signal transduction pathway is being activated, steps are also being taken by the cell to attenuate that acute signal through increased O-glycosylation of many of the same proteins.

These effects of O-GlcNAc on insulin signaling, particularly in rodent models wherein the levels can be effectively and chronically modulated in the physiologic range, are very consistent with the insulin resistance associated with “glucose toxicity” seen in humans. The direct evidence that the HBP may be directly involved in human diabetes, however, is limited and as yet inconclusive, largely because of experimental limitations. In humans, glucose and insulin infusions increase levels of O-GlcNAc in skeletal muscle [75]. Human studies, however, have revealed modest [76] or no [77] effects of glucosamine infusion on insulin resistance. These studies are insufficient to refute the hexosamine hypothesis, however. The glucosamine infusions may not have reached threshold concentrations for sufficient lengths of time or may not have targeted key tissues such as liver or adipose tissue. For example, in the negative study cited earlier, glucosamine infusion into the forearm did not lead to insulin resistance [78] but the study was limited in its sensitivity to detect a modest effect, and moreover the principal target for producing insulin resistance in muscle may be the adipocyte rather than muscle itself [45]. Other evidence in humans does support a role for the HBP. We showed, for example, that insulin resistance in type 2 diabetes was correlated with GFA levels in skeletal muscle [30,54]. Leptin mRNA and protein levels are also correlated with adipocyte UDP-GlcNAc levels that are in turn well correlated with body mass index in humans [79].

**O-GlcNAc signaling as an adaptive response to overnutrition**

The effects are consistent with a role for O-GlcNAc in damping acute hormone- and phosphorylation-mediated signals in situations of chronic nutrient excess. Although discovered in the context of diabetes, the afore-mentioned changes mediated by the HBP can also be viewed as adaptive responses to excess nutrient flux: muscle cells protect themselves from excess glucose fluxes and the excess nutrients are eventually stored as fat. Indeed, if insulin signaling were not dampened and glycogen synthesis were effectively engaged even with overeating, a pound of ingested carbohydrate would result in approximately four pounds of hydrated glycogen in muscle, and it is easy to visualize diets rich in sodas and donuts resulting in the development of glycogen storage diseases. With age and chronic overstimulation of the pathway, however, animals with chronically increased hexosamine flux exhibit several of the maladaptive features of the type 2 diabetes syndrome including obesity, hyperlipidemia, insulin resistance, and β-cell failure.
[44,47,80]. Thus, the HBP can be viewed as both physiologic and pathophysiologic, triggering normal nutrient regulation as well as aspects of “glucose toxicity” and the metabolic syndrome.

**O-GlcNAc and other aspects of glucose toxicity**

The O-GlcNAc pathway also plays a role in hepatic glucose production and insulin secretion, the other key abnormalities of type 2 diabetes. The glycosylation of FOXO1, a key regulator of gluconeogenesis, was mentioned earlier. This glycosylation is mediated, at least in part, through the binding of OGT to O-GlcNAc-modified PGC1α, resulting in the targeting of OGT to binding partners of PGC1α that include the FOXOs [59]. It was also demonstrated that overexpression of O-GlcNAcase in the liver of insulin resistant 

The O-GlcNAc pathway also plays a role in hepatic glucose production and insulin secretion, the other key abnormalities of type 2 diabetes. The glycosylation of FOXO1, a key regulator of gluconeogenesis, was mentioned earlier. This glycosylation is mediated, at least in part, through the binding of OGT to O-GlcNAc-modified PGC1α, resulting in the targeting of OGT to binding partners of PGC1α that include the FOXOs [59].

**Oxidative stress and β-cell failure in diabetes**

Increasing evidence links oxidative stress provoked by hyperglycemia to β-cell damage [91,92]. Interestingly, the latter reference links hexosamine signaling to induction of oxidative stress, providing one potential link between these important processes in glucose toxicity. β-Cells are particularly susceptible to oxidative stress because they contain intrinsically low concentrations of antioxidant enzymes [93]. In addition, β cells are programmed to target most of their glucose to oxidative phosphorylation for signaling of insulin secretion, so that in states of hyperglycemia there will be an added stress of increased production of oxidant species on top of the decreased ability to reduce those species. Additivity of oxidant stress with high concentrations of glucose has been directly demonstrated [94]. These findings have led to multiple studies that demonstrate that transgenic overexpression of antioxidant enzymes (both cytosolic and mitochondrial superoxide dismutase, catalase, and glutathione peroxidase) all protect from β-cell failure in isolated islets, cultured insulin-producing cells, and rodent models [95,96]. As would be predicted from these data, insulin secretory defects and the changes in genes controlling insulin secretion in several experimental models of glucose toxicity have also been prevented by various antioxidant treatments such as N-acetylcysteine, troglitazone, and aminoguanidine [91,97].

**Cellular mechanisms for effects of hyperglycemia on insulin secretion**

Exposure of the β cell to experimental chronic hyperglycemia induces specific desensitization to glucose whereas the insulin response to other secretagogues such as arginine [37], leucine-norbenz [98], and isoproterenol [99] are either preserved or exaggerated. In keeping with this, the insulin content of the islet is not reduced by hyperglycemia to an extent that would account for the degree of glucose-induced desensitization [37]. Such selective desensitization to glucose but not other secretagogues such as arginine is also seen during the early types 1 and 2 diabetes. This desensitization is considered entirely reversible by restoration of normal glucose concentrations [100]. Consistent with these observations, one mechanism by which antioxidants restore normal insulin secretion is direct preservation of mitochondrial function and stimulus-secretion coupling [101].

Insulin content also is diminished by hyperglycemia: in animals and cell lines, hyperglycemia induces a gradual loss of insulin gene expression and gene and protein expression levels of pancreas duodenum homeobox-1 (PDX-1) and MafA
transcription factors, critical regulators of the insulin gene, and these effects are also reversed with antioxidant treatment [91].

**Oxidative stress and insulin resistance**

The association of oxidative stress with insulin resistance has been documented in humans for almost two decades [102]. These effects are mirrored in cell culture models, wherein it has also been shown that antioxidants such as α-lipoic acid protects against the effects of oxidant stress on insulin signaling pathways in cultured adipocytes and myocytes [103]. These studies have prompted trials of antioxidants in humans, wherein short-term improvements in insulin sensitivity are seen [104], although the long-term benefits of antioxidant therapies in type 2 diabetes have been less encouraging. Part of the reason for the latter is likely due to the complexity of oxidant stress on insulin signaling and the function of insulin-responsive tissues. Oxidant stress in diabetes is not only linked to excess glucose and lipid fluxes, but also to ancillary factors such as tissue iron levels [105]. The targets of oxidant stress include mitochondrial respiration rates [106], the proximal insulin-signaling pathway and downstream associated pathways such as p38/MAP kinase and Jnk which are linked to inflammatory signaling [90,107], the AMP-activated protein kinase pathway [108], and numerous others, not to mention more nonspecific effects caused by the nonspecific oxidative damage to DNA, lipids, and proteins. Hence, delivery of the right species of antioxidant to the right tissues at sufficiently sustained levels will be a daunting task.

**Clinical significance of glucose toxicity**

**Type 1 diabetes**

After diagnosis of type 1 diabetes, initiation of insulin therapy induces partial clinical remission in ~30% of the patients during the first year [96]. This honeymoon period is characterized by normoglycemia, recovery of endogenous insulin secretion, and by improved insulin sensitivity [109]. Although correction of several alterations secondary to insulin deficiency, such as increased counterregulatory hormone secretion [110], hyperosmolarity [111], acidosis [112], electrolyte changes [113] and high free fatty acids [114] could contribute to normalization of insulin secretion and sensitivity, reversal of glucose toxicity may also be of importance for the occurrence of remission. In the DCCT, 855 patients had had type 1 diabetes for 1 to 5 years at baseline and of these 303 were considered C-peptide responders (C-peptide level, 0.20 to 0.50 pmol mL⁻¹ after ingestion of a standardized, mixed meal). Responders receiving intensive therapy maintained a higher stimulated C-peptide level and a lower likelihood of becoming nonresponders than did responders receiving conventional therapy (relative risk reduction, 57%) [1].

As discussed earlier, in type 1 patients with long-standing disease, chronic hyperglycemia can be regarded as the key factor responsible for insulin resistance. Therefore and in contrast to the results obtained in patients with type 2 diabetes (see later), insulin sensitivity can be markedly improved and even normalized in type 1 patients by optimizing insulin therapy [20]. The enhanced insulin sensitivity explains why glycemic control can be improved without necessarily having to increase the daily dose [115,116]. Thus, although the loss of insulin secretion is irreversible in type 1 diabetes, insulin sensitivity is amenable to marked modification by alterations in glycemic control.

**Type 2 diabetes**

Data from both several cross-sectional and prospective studies have documented that hyperinsulinemia and insulin resistance both precede and predict the subsequent development of type 2 diabetes. The etiology of insulin resistance is multifactorial and involves familiar/genetic and acquired components. Consequently, one would not expect normalization of insulin sensitivity by improved glycemic control. In keeping with this, insulin resistance has been a uniform finding in patients with type 2 diabetes and has only been partially reversed by, for example, aggressive insulin therapy [7,24,117–119].

Both weight loss [23,120–123], sulfonylureas [123], and insulin therapy [24,25,123–126] improve insulin secretion. Since neither insulin therapy nor weight loss have any direct stimulatory effects on insulin secretion, their effects could be mediated indirectly via diminution of glucose toxicity on β-cell secretion. Indeed, the “extrapancreatic effect” (improved insulin sensitivity) of sulfonylureas has been entirely attributed to amelioration of insulin resistance via lowering of the plasma glucose concentration [127].

In terms of practical clinical care, the fact that the core defects that contribute to hyperglycemia in type 2 diabetes, including excess hepatic glucose production, impaired insulin secretion, and insulin resistance, all improve with control of glycemia means that control of type 2 diabetes should be easier to maintain after a relatively short period of near-normoglycemia. In the largest of such studies 382 Chinese patients were randomly assigned to therapy with insulin or oral hypoglycemic agents [128]. Treatment was stopped after normoglycemia was maintained for 2 weeks after an initial treatment period of ~10 days. Patients were then followed for 1 year on diet and exercise alone. Better glycemic control was achieved with insulin therapy than with oral agents and the remission rates were higher at 1 year in those treated initially with insulin as compared to oral hypoglycemic agents. The acute insulin response was significantly improved by intensive glucose control with insulin. This increase was sustained at 1 year in the insulin groups but significantly declined in patients treated with oral hypoglycemic agents [128]. This study as well as many smaller studies (reviewed in [129]) suggest that short-term intensive insulin therapy early in the course of may offer favorable long-term effects on β-cell function.
Acknowledgments

Supported by grants from the National Institutes of Health (DK 43526 and RR025764, DAM), Veterans Administration Research Service (DAM), and Finnish Academy of Science (HY-J).

References

Glucose toxicity  423

65 Vosseller K, Wells L, Lane MD, Hart GW: Elevated nucleocytoplasmic glycosylation by O-GlcNAc results in insulin resistance.
associated with defects in Akt activation in 3T3-L1 adipocytes. PNAS 2002;99:5313 – 5318.
Glucose toxicity


Monogenic disorders of the β cell

Ali J. Chakera and Andrew T. Hattersley
University of Exeter Medical School, Exeter, UK; Royal Devon and Exeter Hospital Exeter, UK

Key points

- Monogenic β-cell diabetes accounts for 1–2% of diabetes.
- Maturity onset diabetes of the young (MODY) is the commonest form of monogenic β-cell diabetes. MODY can be divided into two clinical groups: mutations in Glucokinase (GCK-MODY), and mutations in the transcription factors—HNF1A and HNF4A (TF-MODY).
- People with GCK-MODY have stable fasting hyperglycemia (5.5–8.0 mmol L\(^{-1}\)) present from birth and do not have significant microvascular complications. Treatment is not required (except in pregnancy).
- TF-MODY diabetes develops in childhood/early adult life and deteriorates over time. Patients are very sensitive to sulfonylurea treatment which is recommended as initial treatment although subsequently insulin treatment may be required. Microvascular complications are common and patients with HNF1A-MODY have an increased cardiovascular risk.
- Maternally inherited diabetes and deafness (MIDD) accounts for approximately 0.5% of diabetes. It is caused by a mitochondrial gene mutation (np3243) and is characterized by maternally inherited diabetes that is initially noninsulin requiring, which usually develops after the onset of sensorineural deafness.
- Neonatal diabetes is a monogenic β-cell diabetes which is diagnosed in the first six months of life. Neonatal diabetes can be permanent or transient, the latter of which remits as an infant, but often reoccurs in later life.
- The commonest causes of neonatal diabetes result from mutations in the \(K_{ATP}\) gene, and excellent glycemic control can be achieved with high-dose sulfonylurea tablets instead of insulin injections.

Introduction

This chapter will concentrate on the monogenic disorders of the β cell that account for 1–2% of diabetes. They are discrete disorders which are a significant cause of diabetes in their own right. Correct molecular diagnosis is important to predict clinical course, explain other associated clinical features, enable genetic counseling, diagnose family members, and most importantly guide appropriate treatment. In addition to this clinical importance, the discovery and study of monogenic disorders has given further insight into the physiology and pathophysiology of the β cell.

In normal health the pancreatic β cell is a finely tuned system that ensures appropriate insulin release in order to maintain homeostasis of blood glucose within a narrow physiologic range. Problems with the β cell can result in diabetes, or more rarely, oversecretion of insulin and hypoglycemia. The most common disease resulting from β-cell dysfunction in humans is type 2 diabetes (T2DM). This is a complex polygenic disorder that is likely to represent a heterogeneous spread of defects.

Monogenic β-cell disorders in key aspects of β-cell function

To understand the pathophysiology of the β cell in monogenic diabetes it is important to outline the process whereby the β cell senses glucose and translates this into an appropriate release of insulin. Figure 28.1 shows a schematic representation of the β cell and the legend describes the process of glucose-insulin secretion coupling. The sites of monogenic defects (indicated on the figure) that cause diabetes or hypoglycemia are as follows:

- GLUT2: GLUT2 mutations cause Fanconi–Bickel syndrome. The phenotype is predominantly due to defects in hepatic GLUT2 causing glycogen storage disease, rather than pancreatic GLUT2. This will not be discussed in this chapter.
- GCK: Loss of function mutations in GCK cause a rise in fasting glucose with maintained homeostasis, referred to as glucokinase—maturity onset diabetes of the young (GCK-MODY).
- Glycolysis: Defects in glycolysis are described in transcription factor MODY (TF-MODY) due to mutations in Hepatocyte nuclear factor (HNF)-1 alpha (HNF1A) and HNF4A.
Glucose enters the β cell via the GLUT2 transporter where it is phosphorylated to glucose-6-phosphate (G6P) by the enzyme glucokinase (GCK). G6P is metabolized via the glycolytic pathway to pyruvate. Pyruvate enters the mitochondria and is metabolized via the citric acid cycle to produce a rise in intracellular ATP and a fall in ADP. This increase in ATP/ADP acts to close ATP-dependent K channels (Kir6.2/SUR1) allowing a rise in the membrane resting potential, which triggers a voltage-gated calcium channel to open, allowing calcium influx. It is this calcium influx that is the trigger for insulin exocytosis.

- Mitochondria: Mitochondrial defects are described in: maternally inherited diabetes and deafness (MIDD).
- \( \text{K}_{\text{ATP}} \) channel: Mutations in \( \text{SUR1} \) and \( \text{Kir}6.2 \) (components of the \( \text{K}_{\text{ATP}} \) channel) cause both permanent and temporary neonatal diabetes and have been shown to cause hyperinsulinemia of infancy.
- Endoplasmic reticulum: Wolfram syndrome (\( \text{WFS-1} \)) and Wolcott–Rallison syndrome (\( \text{EIF2AK3} \)) genes encode proteins that localize to the endoplasmic reticulum.
- Insulin: Insulin gene mutations causing mutant insulin which misfold and cannot be secreted resulting in neonatal diabetes.

**Clinical categories of monogenic β-cell disorders**

This chapter will discuss the three most important clinical categories of monogenic disorders of the β cell: (1) Maturity-onset diabetes of the young, (2) Neonatal diabetes, and (3) Diabetes with marked extrapancreatic features (see Figure 28.2).

**Maturity-onset diabetes of the young**

Maturity-onset diabetes of the young (MODY) refers to group of monogenic subtypes of diabetes characterized by young onset (usually before 25 years) of non-insulin-dependent diabetes, β-cell dysfunction that is inherited as an autosomal dominant trait. Mutations in at least 10 genes have been found to cause MODY; (see Figure 28.3) the common subtypes are due to mutations in \( \text{GCK} \), Hepatocyte nuclear factor 1-alpha (\( \text{HNF1A} \)), \( \text{HNF4A} \), and \( \text{HNF1B} \). The definition of the underlying genes has resulted in the recognition of distinct clinical and physiologic subgroups of MODY with varying clinical course, prognosis, and treatment requirements. Classifications of diabetes by the ADA and the WHO recognize these discrete subtypes.
Diagnosis of MODY
Table 28.2 shows the key differences between type 1 and type 2 diabetes and MODY in young adults. Over 80% of people with MODY will initially be misdiagnosed with type 1 or type 2 MODY [1]. MODY subjects tend to be (but are not exclusively) non-obese, lack features of the insulin resistance syndrome, and be GAD and IA-2 antibody negative. The strict classic criteria for MODY: (i) age of onset diabetes <25, (ii) non-insulin-dependent diabetes, (iii) family history of diabetes (due to autosomal dominant inheritance), have been successfully applied to genetic studies. However, they have poor sensitivity, missing 50% of patients with a diagnosis of MODY [1]. Recently, an online clinical prediction model has been developed for MODY (a MODY calculator: www.diabetesgenes.org/content/mody-probability-calculator). This allows clinicians to identify those patients that are most likely to benefit from genetic testing based on a pretest probability of MODY, taking into account a patient’s age of diagnosis, treatment, BMI, and family history [1]. The importance of a genetic diagnosis of MODY is that these patients often do not need insulin treatment and an appropriate switch of therapy can lead to an improvement in well-being and glucose control.

Prevalence of MODY subtypes
MODY accounts for approximately 1% of non-insulin-dependent diabetes in Europe [1]. Large national collections allow assessment of the relative prevalence of the different subgroups of MODY as shown in Figure 28.3 [2]. The relative prevalence varies, by country, according to how patients are identified. As glucokinase mutations are common (≥0.1% population) but asymptomatic, their frequency will be higher in populations where there is wider screening for and follow-up of abnormal glucose levels in children and young adults.

The MODY phenotypes
The MODY phenotypes can essentially be divided into two distinct clinical groups: glucokinase MODY (GCK-MODY) and transcription factor MODY (TF-MODY). The phenotypes in these two groups are completely different (see Table 28.1), and it is confusing that they both fall under the same group name of MODY. GCK-MODY is characterized by stable fasting hyperglycemia that is present from birth, nonprogressive, does not require treatment (except in pregnancy), and is not associated with serious vascular complications. TF-MODY diabetes, on the other hand, is not present at birth but usually develops in adolescents/young adults. The two commonest forms are due to mutations in HNF1A and HNF4A. They are progressive with increasing treatment requirements and are associated with the full range of diabetes-associated microvascular complications [3].

Glucokinase MODY
Glucokinase catalyzes the phosphorylation of glucose to glucose-6-phosphate. Its unique kinetic properties result in the rate of glucose phosphorylation being proportional to the physiologic glucose concentration. This means that the β cell and hepatocyte can respond appropriately to fluctuations in the degree of glycemia. In the pancreatic β cell, glucokinase acts as “the glucose sensor” [4] ensuring insulin release is appropriate to the blood glucose concentration.

Molecular genetics of glucokinase MODY
Coding or intron/exon boundary mutations were first found in the glucokinase gene in French and English MODY pedigrees in early 1992 [5,6]. Over 200 different heterozygous loss-of-function mutations have now been described, all causing a similar clinical picture. Expression studies show that the mutations in MODY families may alter the enzyme’s affinity or activity for glucose metabolism or phosphorylation or alter structure or stability [7–10]. Homozygous loss-of-function glucokinase mutations are a rare cause of insulin-requiring diabetes presenting in the neonatal period [11]. Gain-of-function mutations cause congenital hyperinsulinism [12].

Phenotype of glucokinase MODY
Subjects with GCK-MODY have fasting hyperglycemia that is present from birth and shows very little deterioration with age [13]. Byrne and colleagues have shown, using graded glucose infusions, that heterozygous mutations in GCK cause a right shift in the insulin response curve leading to a re-setting of the fasting glucose set point [14]. So the majority of GCK-MODY subjects have high fasting plasma glucose within a tight range (5.5–8 mmol L⁻¹). There is maintained glucose homeostasis as shown by the small increment in plasma glucose at two hours during an oral glucose tolerance test, albeit at a higher set point than in normal subjects [15]. The HbA1c is raised in people with GCK-MODY with a median value of 50 mmol mol⁻¹ that, as in normal subjects, deteriorates with age [13]; over half of cases exceed the diagnostic criteria for diabetes. HbA1c values above 64 mmol mol⁻¹ would suggest an alternative diagnosis and marked worsening of the glycemia suggests that the patient has developed type 1 or type 2 diabetes in addition to their GCK mutation. Subjects with GCK-MODY do not get
### Table 28.1 Comparison of the clinical characteristics of glucokinase and transcription factor MODY

<table>
<thead>
<tr>
<th></th>
<th>Glucokinase MODY</th>
<th>Transcription factor MODY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genetic etiology</td>
<td>Loss-of function mutations in glucokinase gene</td>
<td>Loss-of function mutations in HNF1A and HNF4A genes</td>
</tr>
<tr>
<td></td>
<td>Autosomal dominant</td>
<td>Autosomal dominant</td>
</tr>
<tr>
<td>Onset of hyperglycemia</td>
<td>From birth (though asymptomatic)</td>
<td>Second / third decade</td>
</tr>
<tr>
<td>Presentation</td>
<td>Usually asymptomatic, detected by screening or on routine testing</td>
<td>Usually symptomatic</td>
</tr>
<tr>
<td>Details of hyperglycemia</td>
<td>Mild stable hyperglycemia (fasting glucose 5.5–8 mmol L(^{-1}))</td>
<td>May be severe (FPG frequently &gt;14 mmol L(^{-1}) of treatment)</td>
</tr>
<tr>
<td></td>
<td>Minimal increase in glycemia with age (parallel to normals)</td>
<td>Progressive deterioration of glycemia with age</td>
</tr>
<tr>
<td></td>
<td>HbA1c 40–60 mmol L(^{-1}) (5.7–7.7%)</td>
<td>HbA1c variable depending on control, may be high (as with type 2 diabetes)</td>
</tr>
<tr>
<td>Pattern in an oral glucose tolerance test</td>
<td>Fasting: &gt;5.5 mmol L(^{-1})</td>
<td>Fasting: often &lt;5.5 mmol L(^{-1})</td>
</tr>
<tr>
<td></td>
<td>120 min: increment usually &lt;4.4 mmol L(^{-1})</td>
<td>120 min: increment usually &gt;4.4 mmol L(^{-1})</td>
</tr>
<tr>
<td>Microvascular complications</td>
<td>No significant microvascular complications</td>
<td>Frequent</td>
</tr>
<tr>
<td>Macrovascular complications</td>
<td>As normal population</td>
<td>HNF1A:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– Raised cardiovascular risk</td>
</tr>
<tr>
<td>Pathophysiology</td>
<td>Glucose sensing defect</td>
<td>Insulin secretion maintained at normal glucose values but not increased in hyperglycemia</td>
</tr>
<tr>
<td>Extrapancreatic manifestations</td>
<td>None</td>
<td>HNF1A:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– Low renal glucose threshold</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– Raised HDL</td>
</tr>
<tr>
<td>Pregnancy</td>
<td>Fetal inheritance leads to a 550 g decrease in birth weight</td>
<td>HNF1A:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– No fetal effect on birth weight has been observed</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HNF4A:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– Possible fetal hypoglycemia</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– Increased risk of fetal macrosomia (fetal inheritance leads to a 800 g increase in birth weight)</td>
</tr>
<tr>
<td>Treatment</td>
<td>None, unless co-existent type 1 or type 2 diabetes</td>
<td>Sensitive to sulfonylurea treatment</td>
</tr>
<tr>
<td></td>
<td></td>
<td>May progress to require insulin</td>
</tr>
</tbody>
</table>

### Table 28.2 Differentiating β cell monogenic diabetes from type 1 and type 2 diabetes in young adults

<table>
<thead>
<tr>
<th>Features</th>
<th>Type 1 diabetes</th>
<th>Young-onset type 2 diabetes</th>
<th>GCK-MODY</th>
<th>TF-MODY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insulin dependent</td>
<td>Yes—after</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>honeymoon</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parent affected</td>
<td>2–4%</td>
<td>60%</td>
<td>80%—but may not be diagnosed</td>
<td>80%</td>
</tr>
<tr>
<td>Usual age of onset</td>
<td>2 years—young adult</td>
<td>Adolescent and young adult</td>
<td>Birth (diagnosed at any age)</td>
<td>Adolescent and young adult</td>
</tr>
<tr>
<td>Obesity</td>
<td>Pop freq</td>
<td>&gt;90%</td>
<td>Pop freq</td>
<td>Pop freq</td>
</tr>
<tr>
<td>Acanthosis nigricans</td>
<td>No</td>
<td>&gt;75%</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Glycemia</td>
<td>High</td>
<td>Variable—progressive</td>
<td>Mild stable</td>
<td>Variable—progressive</td>
</tr>
<tr>
<td>β-cell auto-antibodies</td>
<td>Positive in 80% of patients at diagnosis</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

GCK, glucokinase; HNF1A, hepatocyte nuclear factor 1A (HNF4A is similar); pop freq, the frequency seen in the general population.
significant microvascular complications [13]. The incidence of macrovascular complications appears similar to the background population.

Diagnosis
As GCK mutations cause mild asymptomatic hyperglycemia most cases of GCK-MODY are diagnosed following incidental testing of blood glucose, or following routine screening in pregnancy. A correct diagnosis is important in young people who might otherwise be thought to have type 1 diabetes. Persistent, stable mild hyperglycemia, lack of β-cell autoantibodies and mild hyperglycemia in a parent (when tested) all support a diagnosis of GCK-MODY. Differentiating GCK-MODY from T2DM can be more difficult; however, supporting clinical evidence for GCK-MODY includes lack of obesity or features of insulin resistance, or results obtained from an oral glucose tolerance test, when the glucose increment at 2 hours is usually <3 mmol L\(^{-1}\) [15]. Fasting plasma glucose below 5.5 mmol L\(^{-1}\) is rare in GCK-MODY outside pregnancy.

Treatment
Patients with GCK-MODY do not usually require treatment except in pregnancy (see later) [16]. If a patient is being treated with insulin or other hypoglycemic agents, it is usually possible to stop treatment when a diagnosis of GCK-MODY is made. However, if the HbA1c has ever exceeded 64 mmol L\(^{-1}\), this should be done with caution, as it is possible for someone to have both a GCK mutation and type 1 or type 2 diabetes.

GCK-MODY and pregnancy
Women with glucokinase mutations are frequently first found to have hyperglycemia during screening in pregnancy. GCK-MODY represents about 3% of patients with gestational diabetes [17]. Correctly identifying GCK-MODY is important because these women have a different clinical course both within and outside pregnancy compared to “standard” gestational diabetes.

The aim of treatment of diabetes in pregnancy is to minimize the risk of fetal and maternal complications. This is usually achieved by strict control of the mother’s blood glucose to prevent fetal macrosomia. However, in GCK-MODY pregnancies, the need to reduce maternal blood glucose is not absolute. This is because the fetus has a 50% chance of inheriting the GCK mutation from the mother, and the presence of the GCK mutation in the fetus influences the fetus’ sensing of maternal glucose. Therefore, if the fetus does not inherit the GCK mutation it will respond to maternal hyperglycemia by excess insulin production and therefore excess growth, whereas if the fetus does inherit the GCK mutation it will sense the maternal hyperglycemia as normal, produce normal amounts of insulin and have normal growth (see Figure 28.4) [18]. In this latter setting, aggressively lowering maternal glycemia may adversely affect fetal growth [19]. As it is not usually possible to determine the fetal genotype, the decision on whether to treat the mother should be based on fetal growth. If the abdominal circumference is greater than the 75th centile insulin may be used but early delivery is the most successful strategy (see Figure 28.5).

**Figure 28.4** Corrected mean birth weight for offspring of parents with glucokinase-MODY by category of maternal and fetal mutation status.
A mother with a GCK mutation will need no pharmacologic treatment postpartum even if she has received very high insulin doses in pregnancy. This is in contrast to subjects with nonglucokinase gestational diabetes where deterioration to T2DM is usually seen over the following 10 years.

**Transcription factor MODY**

The majority of MODY genes encode transcription factors (TF). Transcription factors have an important role in regulating the expression of genes. They have a critical role in the variations in gene expression in the different stages of embryological development, and in different tissues. Transcription factors can also regulate expression of other transcription factors, thus establishing a complex regulatory network controlling gene expression.

**Molecular genetics of TF-MODY**

*Hepatocyte nuclear factor-1α (HNF1A)* was the first identified TF gene to cause MODY [20]. *HNF1A* is found on the long arm of chromosome 12 (12q), and mutations in *HNF1A* are the most common cause of MODY in the UK and most of Europe [2]. Nearly 200 mutations have been found scattered throughout the gene and consist of frameshift, missense, nonsense, and splice site mutations.

Transcription factor mutations alter insulin secretion in the mature β cell as well as altering β-cell development, proliferation, and cell death. Mutations in the hepatic nuclear factors appear to alter levels of proteins critical in metabolism including the GLUT2 glucose transporter and key enzymes in the mitochondrial metabolism of glucose [21,22].

Mutations in *HNF4A* are a much less common cause of MODY than *HNF1A* mutations, accounting for only 3% of MODY in the UK. Over 40 mutations have been described with the majority in the first four exons of the gene [23].

Mutations in the *HNF1B* gene cause renal cysts and diabetes syndrome and will be discussed under the heading “Diabetes with extra-pancreatic features” later in this chapter.

Other transcription factor mutations causing autosomal dominant β-cell diabetes have been identified in the following genes *PDX-1(IPF-1)*, *NEUROD1*, *PAX 4*, and *KLF11* but all are very rare and will not be described in this chapter [24–29].

**HNF1A and HNF4A**

From mouse models and *in vitro* studies in patients with TF-MODY, the development of diabetes and its deterioration with time is likely to be due to a metabolic defect in β-cell glycolysis and progressive loss of β-cell mass [30,31].

The pancreatic phenotypes in HNF1A-MODY and HNF4A-MODY have been the most studied of the TF-MODYs. It is striking that the different TFs causing MODY can have such differing nonpancreatic phenotypes (e.g. abnormal lipids, renal cysts) but have essentially the same pancreatic phenotype. The TFs causing MODY all form part of a regulatory network. Therefore, in the pancreas (but not in other tissues where an alternative HNF4A promoter is not used) there is a positive feedback loop established between HNF1A and HNF4A. It has been hypothesized that this feedback loop acts as a bistable switch that is only stable fully on or fully off. Therefore loss of function of HNF1A or HNF4A directly, or indirectly via *HNF1B* and *PDX-1 (IPF-1)* mutations, will turn the switch to off thereby causing the same pancreatic phenotype whatever TF is mutated in this network [32]. The lack of this positive feedback switch in other tissues means that the nonpancreatic phenotype is determined by the role of the specific TF involved.

**Phenotype of HNF1A and HNF4A**

Heterozygous transcription factor mutations cause autosomal dominant diabetes presenting in adolescence or early adulthood resulting from progressive failure of insulin secretion. While diabetes is similar in HNF1A and HNF4A mutation carriers as a result of a common pattern of β-cell dysfunction, a number of differences in extrapancreatic features occur.

**Common diabetes phenotype of HNF1A and HNF4A**

The pancreatic phenotype in TF-MODY is characterized by:

- **β-Cell dysfunction.** The β-cell defect with reduced insulin secretion in response to stimulus has been shown in HNF1A- [33] and HNF4A-MODY [34] using graded glucose infusions. There is also marked loss of the first phase of insulin release in diabetic HNF1A subjects [3].

- **Progressive deterioration with time.** HNF1A-MODY subjects show a progressive deterioration in fasting plasma glucose with age [3]. This deterioration explains the increasing treatment requirements with time. Mouse studies [31,35] suggest that the deterioration is due to a decrease in β-cell mass with time. β-Cell loss is not auto-immune mediated as pancreatic autoantibodies are found in the same proportion as the nondiabetic general population.

In people with HNF1A- and HNF4A-MODY, diabetes usually develops in adolescence or early adult life. The youngest age of onset of diabetes in HNF1A-MODY reported is 4 years, with 63% of patients having developed diabetes by the age of 25; 79% by 35 years and 96% by 55 years. The mean age of diagnosis is similar in HNF1A- (20 years) and HNF4A-MODY (23 years) [36]. The age of diagnosis is in part related to the genetic mutations. In HNF1A there is evidence of imprinting...
in utero, as the offspring inheriting the mutation of mothers with HNF1A diabetes during pregnancy have a younger age of onset than those mothers who develop the diabetes after their pregnancy.

Nondiabetes features of HNF1A and HNF4A

**HNF1A**

HNF1A-MODY patients are characterized by a low renal threshold for glucose. Hence glycosuria can occur at relatively mild levels of blood glucose (<8 mmol L⁻¹). Renal glycosuria has been shown to precede the development of diabetes and makes the screening of at-risk children by urinary screening for glycosuria an appropriate and sensitive screening test.

Despite patients with HNF1A-MODY having elevated levels of HDL, they also have an increased risk of cardiovascular disease compared with patients with T1DM. When compared with family members without HNF1A-MODY, those with a HNF1A mutation die at a younger age with a hazard ratio of 2.6 for cardiovascular death [37]. Frequency of microvascular complications is similar to that seen in type 1 and type 2 diabetes and relates to degree of glycemic control [38].

**HNF4A**

The offspring of HNF4A mutation-carrying mothers and fathers are at risk of marked macrosomia. There is on average an 800 g increase in birth weight compared with nonmutation-carrying siblings [39]. There is also an increased risk of hypoglycemia in affected neonates. These features appear to relate to increased insulin secretion in utero and in early infancy which evolves into reduced insulin secretion and diabetes in later life [39].

HNF4A-MODY patients have reduced concentrations of the apolipoproteins apoAII, apoCIII, and possibly apoB which are not seen in patients with T2DM [40,41]. Triglyceride levels are reduced in patients with HNF4A-MODY and probably reflect reduced lipoprotein lipase activity as a result of the reduced apolipoproteins [40].

Diagnosis of HNF1A and HNF4A

MODY should be considered in any person who develops diabetes at a young age. As genetic testing is currently too expensive for widespread use, the key diagnostic challenge is identifying TF-MODY in cases that would typically be labeled type 1 or type 2 diabetes. The key discriminatory clinical features are outlined in Table 28.3. As most patients with TF-MODY are diagnosed when young and do not typically have features associated with insulin resistance, they are most often labeled T1DM. Key suspicious features that would lead to molecular testing for TF-MODY include being β-cell autoantibody negative with a first-degree family relative with diabetes; no ketosis in the absence of insulin treatment; persistent insulin production (C-peptide) several years after diagnosis (post honeymoon) [42].

HNF1A- and HNF4A-MODY should be suspected in patients who have been labeled as having T2DM when there is a personal and family history of diabetes diagnosed at a young age (<25) and where there are no signs associated with insulin resistance (obesity, acanthosis nigricans).

Once MODY is suspected, a molecular genetic diagnosis should be sought. Clinical features can be used to target the gene analyzed. Although as HNF1A is more common than HNF4A a strategy in diagnostic testing is to test for HNF1A where there is reasonable clinical suspicion and only test for HNF4A in those who are negative for HNF1A.

In family members of those with an HNF1A or HNF4A mutation all family members with a pre-existing diagnosis of diabetes should have molecular genetic testing for the known mutation. In those without diabetes, it is usually sufficient to screen for diabetes using fasting plasma glucose. In young HNF1A mutation carriers (<20 years), the fasting plasma glucose can be normal, so testing urine for glycosuria can screen those where an oral glucose tolerance test is needed to confirm a diagnosis of diabetes; there is often a diabetic 2-h value, with a large increment between the fasting and 2-h plasma glucose [15].

Treatment of HNF1A and HNF4A

Most patients with a new diagnosis of HNF1A or HNF4A can be managed with dietary advice before requiring oral agents. When commencing oral agents the sensitivity to sulfonylureas means that these are the treatment of choice [43]. Even very low sulfonylurea doses may cause hypoglycemia. The starting dose should therefore be low—we use a starting dose of 40 mg gliclazide daily in adults. If there is hypoglycemia with low doses of sulfonylurea a short-acting agent such as nateglinide may be appropriate [44].

In those patients on treatment for a pre-existing diagnosis of type 1 or type 2 diabetes, the aim should be to switch to treatment with a sulfonylurea (if this is not already the case). Glycemic control and well-being with sulfonylureas is often better than on insulin and the fasting glucose-lowering effect is four times greater than that seen in T2DM [43]. Sulfonylurea treatment is successful in the majority of patients although a DP4 inhibitor or insulin therapy may be required as diabetes progresses [45].

In view of the increased risk of cardiovascular disease in HNF1A, statin therapy should be considered for all patients aged over 40 years; as in T1DM.

**HNF1A and HNF4A in pregnancy**

Evidence for the management of HNF1A and HNF4A in pregnancy is limited to case experience. As with all pregestational diabetes prenatal counseling is sensible to discuss and optimize treatment.

Sulfonylureas have been used successfully in HNF1A and HNF4A pregnancies with well-controlled diabetes. Given the greater clinical experience and pregnancy safety data with glibenclamide (glyburide) we recommend that women switch to this sulfonylurea prior to conception. If control remains suboptimal insulin treatment will be needed in addition.
Table 28.3 Causes and classification of neonatal diabetes

<table>
<thead>
<tr>
<th>Pancreatic pathophysiology</th>
<th>Protein, chromosome, or gene affected</th>
<th>Prevalence</th>
<th>Inheritance</th>
<th>Features in addition to neonatal diabetes and low birth weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reduced β-cell function</td>
<td>KATP channel (KCNJ11 and ABCC8)</td>
<td>35% of PNDM and 25% of TNDM</td>
<td>85% spontaneous. Remainder autosomal dominant or recessive</td>
<td>Developmental delay and epilepsy–genotype–phenotype correlation. (DEND and iDEND). Sulfonylurea responsive</td>
</tr>
<tr>
<td></td>
<td>Chromosome 6q24</td>
<td>60% of TNDM</td>
<td>Variable</td>
<td>Macroglossia and umbilical hernia</td>
</tr>
<tr>
<td></td>
<td>GCK (homozygous for mutation)</td>
<td>Rare cause of PNDM</td>
<td>Autosomal recessive</td>
<td>Both parents have GCK-MODY</td>
</tr>
<tr>
<td></td>
<td>GLIS3</td>
<td>Rare cause of PNDM</td>
<td>Autosomal recessive</td>
<td>Congenital hypothyroidism, glaucoma, liver fibrosis, and cystic kidney disease</td>
</tr>
<tr>
<td></td>
<td>SLC2A2</td>
<td>Rare cause of TNDM</td>
<td>Autosomal dominant or recessive</td>
<td>Thiamine-responsive megaloblastic anemia: hypergalactosemia, hepatic failure, neurologic deficit</td>
</tr>
<tr>
<td>Reduced pancreatic mass</td>
<td>GATA6</td>
<td>Rare cause of pancreatic aplasia</td>
<td>Autosomal dominant</td>
<td>Pancreatic agenesis and hypoplasia; cardiac malformations; developmental delay; gastrointestinal hernias</td>
</tr>
<tr>
<td></td>
<td>PTF1A</td>
<td>Rare cause of pancreatic aplasia</td>
<td>Autosomal recessive</td>
<td>Pancreatic agenesis; cerebellar agenesis; developmental delay</td>
</tr>
<tr>
<td></td>
<td>PDX1</td>
<td>Rare cause of pancreatic aplasia</td>
<td>Autosomal recessive</td>
<td>Pancreatic agenesis</td>
</tr>
<tr>
<td></td>
<td>HNF1B</td>
<td>Rare cause of TNDM</td>
<td>Autosomal dominant</td>
<td>Exocrine pancreas insufficiency and renal cysts</td>
</tr>
<tr>
<td>Increased β-cell destruction</td>
<td>INS</td>
<td>12% of PNDM</td>
<td>Autosomal dominant</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>EIF2AK3</td>
<td>30% of PNDM in consanguineous parents</td>
<td>Autosomal recessive</td>
<td>Wolcott–Rallison syndrome: spondylo-epiphyseal dysplasia, renal failure, recurrent hepatitis and mental retardation</td>
</tr>
<tr>
<td></td>
<td>FOXP3</td>
<td>Rare cause of PNDM</td>
<td>X-linked</td>
<td>Immune dysregulation; thyroid autoimmunity; intractable diarrhea, eczematous skin rash, and elevated IgE</td>
</tr>
</tbody>
</table>

In pregnancies where either parent has an HNF4A mutation, the offspring is at increased risk of macrosomia and neonatal hypoglycemia if they inherit the mutation (a 50% chance) [46]. Given this, pregnancies need close monitoring with: very tight glucose control in mothers with diabetes; frequent fetal ultrasounds to monitor growth; early (even preterm) planned delivery if evidence of macrosomia; and monitoring of the newborn for hypoglycemia with consideration of diazoxide treatment, if hypoglycemia persists, according to published guidelines [47].

Unlike HNF4A, in pregnant women with HNF1A-MODY, fetal inheritance of the mutation does not affect outcome, so treatment should focus on controlling maternal hyperglycemia to reduce the risk of macrosomia.

Neonatal diabetes: diabetes diagnosed within 6 months of life

Neonatal diabetes (NDM) is defined as monogenic β-cell diabetes which is diagnosed in the first six months of life. It is rare, affecting one in 200,000 live births [48]. The evidence that a diagnosis before 6 months is the cut-off between monogenic
neonatal diabetes rather than polygenic T1DM comes from studies of high-risk type 1 HLA, antibodies, birth weight (reduced before 6 months suggests a reduced insulin secretion in utero) and monogenic genetic studies [49–52].

**Clinical features**

Patients may be detected when glucose is tested in the first few days of life in low-birth weight babies or may present later when the patient is severely unwell with ketoacidosis. Patients are severely insulin deficient and C-peptide is usually not detectable. Approximately 30–40% of cases remit spontaneously and are referred to as transient neonatal diabetes mellitus (TNDM) as opposed to permanent neonatal diabetes mellitus (PNDM) where diabetes persists. TNDM often recurs between the first and third decades of life [53]. Thirteen genes involved in β-cell development, survival, or function have been implicated in the development of neonatal diabetes. Table 28.3 summarizes the known genetic causes of neonatal diabetes.

**Permanent neonatal diabetes**

**Molecular genetics**

Over a third of NDM is caused by mutations in the genes *KCNJ11* and *ABCC8*. These encode the Kir6.2 and SUR1 subunits, respectively, of the β cell ATP-sensitive potassium channel (K$_{ATP}$ channel) [54–57]. Normally, the K$_{ATP}$ channel closes in response to raised intracellular ATP levels that occur as blood glucose rises. Channel closure causes depolarization of the β-cell membrane which results in insulin secretion. Activating mutations in *KCNJ11* and *ABCC8* prevent the K$_{ATP}$ channel closing, in response to increased ATP, preventing insulin release [58].

The K$_{ATP}$ channel is also present in the brain, nerves, and muscles. In those with more functionally severe *KCNJ11* mutations (~20% patients), and occasional patients with *ABCC8* mutations, there is associated neurodevelopmental delay [52,54,55,58].

The majority (80%) of PNDM resulting from K$_{ATP}$ channel mutations arise spontaneously from de novo heterozygous mutations, with the remainder inherited mainly in an autosomal dominant pattern. About 25–40% of neonatal diabetes resulting from *ABCC8* mutations is inherited in an autosomal recessive fashion [57].

Heterozygous mutations in the insulin gene (INS) have been identified in 10% of cases of neonatal diabetes [59]. *EIF2AK3* recessive mutations cause Wolcott–Rallison syndrome and are the commonest cause of neonatal diabetes when the parents are consanguineous and account for 24% of these cases [60]. A number of other genetic causes have been found which all are relatively rare as outlined in Table 28.3 [61].

**Clinical features of the common subtypes of permanent neonatal diabetes**

Neonatal diabetes, caused by *KCNJ11*, *ABCC8*, and INS mutations, typically present within the first six months of life (median 6–10 weeks). Infants are unwell, failing to thrive, with marked hyperglycemia often accompanied by ketosis. C-peptide is usually undetectable and pancreatic autoantibodies are negative [54]. Birth weight is low (~2.5 kg) as a result of reduced insulin secretion in utero with consequent decreased insulin-mediated growth.

About 20% of patients with PNDM and *KCNJ11* mutations have neurologic features, the commonest being developmental delay, sometimes with muscle weakness and/or epilepsy. The most severe form where neonatal diabetes is accompanied by developmental delay and epilepsy has been named “DEND” (developmental delay, epilepsy, and neonatal diabetes). “Intermediate DEND” (iDEND) refers to neonatal diabetes with less severe developmental delay. The severity of the clinical condition relates closely to the underlying mutation and its effect on K$_{ATP}$ channel ATP sensitivity [58,62].

Neonatal diabetes caused by *ABCC8* mutations leads to TNDM almost as often as PNDM and rarely has associated neurodevelopmental features [55–57]. Neonatal diabetes due to INS mutations causes PNDM but patients do not have extrapancreatic features [63].

Patients with Wolcott–Rallison syndrome typically develop diabetes before six months, and skeletal dysplasia occurs within the first two years of life. Other manifestations vary between patients in their nature and severity and include frequent episodes of acute liver failure, renal dysfunction, exocrine pancreas insufficiency, intellectual deficit, hypothyroidism, neutropenia, and recurrent infections [60,64].

**Treatment of permanent neonatal diabetes**

At diagnosis of neonatal diabetes, all patients should be commenced on insulin treatment to stabilize blood sugars. All patients with neonatal diabetes should be referred for urgent molecular genetic testing, results of which should be available within a week. This is because over a third of patients will have a K$_{ATP}$–channel mutation (*KCNJ11* or *ABCC8*) the majority of whom can successfully transfer from insulin to sulfonylurea therapy; 90% of those with *KCNJ11* mutations are able to discontinue insulin. There are usually significant improvements in glycemic control without an increase in hypoglycemia [56,65]. Sulfonylureas work by closing the β cell K$_{ATP}$ channel by an ATP-independent route thus allowing β-cell depolarization and insulin release [65]. Glibenclamide has been used in the majority of NDM cases. It was initially used as it is nonselective and widely available; however there is evidence that it may be more effective than other sulfonylurea agents [66]. The doses needed are higher than those needed for the treatment of T2DM: a median dose of 0.45 mg kg$^{-1}$ per day is required with doses up to 1.5 mg kg$^{-1}$ d$^{-1}$ needed in some cases [65,67]. Sulfonylurea therapy may result in some improvement in neurologic features even when they are commenced in adulthood [66,68,69].

All other causes of neonatal diabetes need treatment with insulin. A molecular genetic diagnosis is still important as it will inform on extrapancreatic clinical features and allow genetic counseling [70].
Table 28.4 Characteristics of diabetes with extrapancreatic features

<table>
<thead>
<tr>
<th>Syndrome</th>
<th>Gene defect</th>
<th>Extrapancreatic clinical features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Renal cysts and diabetes (RCAD)</td>
<td>HNF1B</td>
<td>Renal cysts or renal development disorders, Short stature, Genito-urinary abnormalities, Hyperuricemia and gout, Hypermagnesemia, Abnormal liver function tests</td>
</tr>
<tr>
<td>Mitochondrial diabetes and deafness (MIDD)</td>
<td>m3243A&gt;G</td>
<td>Maternal inheritance, Neurosensory deafness, Pigmentary retinal dystrophy, Nephropathy, Neuromuscular features</td>
</tr>
<tr>
<td>Wolfram syndrome (also known as DIDMOAD)</td>
<td>WFS1</td>
<td>Diabetes Insipidus, Diabetes Mellitus, Optic Atrophy and Deafness, Degenerative neurologic disease, Dilated nephropathy</td>
</tr>
</tbody>
</table>

Transient neonatal diabetes

Molecular genetics

The genetic etiology of more than 80% of transient neonatal diabetes has been established. The commonest cause is abnormalities in the q24 region of chromosome 6 (6q24) affecting imprinted genes [61,71]. Genetic imprinting occurs when only the maternal or paternal inherited allele of a gene is expressed. In TNDM, paternal uniparental disomy, paternal duplication of 6q24 or abnormal methylation of the maternal copy of the chromosome causes overexpression of the paternal copies of the genes PLAGL1 (also known as ZAC) and HYMAI [71,72]. Paternal duplication of 6q24 can be inherited; therefore this abnormality causes the majority of inherited TNDM cases. Uniparental disomy causes sporadic TNDM; cases resulting from abnormal methylation of the maternal copy of chromosome 6 may be sporadic or inherited [71,72]. The majority of TNDM not associated with 6q24 abnormalities are caused by mutations in KCNJ11 and ABCC8 [52,53,56,73–76].

Clinical features of transient neonatal diabetes

6q24 diabetes usually presents in the first week of life. Neonates often have severe hyperglycemia and dehydration but are usually not ketotic [71]. Islet cell antibodies are usually negative and C-peptide is low or negligible [71]. Low birth weight is common (mean birth weight 2 kg), and there may be associated macroglossia and/or umbilical hernia. Insulin treatment is required for a median of 12 weeks before the patient goes into remission. Diabetes recurs later in life in 50–60% of patients as a result of pancreatic β-cell dysfunction. The average age of recurrence is 14 years. In some cases hyperglycemia may be intermittent and seen only at times of stress [71,77].

Where TNDM is caused by KCNJ11 and ABCC8 mutations, diabetes tends to present later (median 4 weeks), takes longer to remit, and is associated with less intrauterine growth restriction (median birth weight 2.6 kg). Remission occurs in about 50% of patients and tends to occur in the first decade (median 4.7 years (3 – 15)) [53].

Management of transient neonatal diabetes

Insulin is required at diagnosis in the neonatal period; however, on relapse in later life, treatment requirements vary: diet; oral hypoglycemic agents or insulin [77]. In TNDM cases that result from KCNJ11 or ABCC8 mutations, infants may be switched from insulin to sulfonylurea treatment [53,56]. The sulfonylurea doses required are lower than for infants with K_ATP-channel mutations causing PNDM and care should be given to monitor for hypoglycemia.

Genetic counseling depends on the underlying genetic etiology. Cases caused by uniparental disomy are sporadic and therefore have low risk of occurrence in either siblings or offspring of the affected child. Methylation defects often result from homozygous mutations in the transcription factor gene ZFP57 and therefore may be inherited in an autosomal recessive manner [72]. Offspring of males with 6q24 duplication have a 50% chance of developing TNDM whereas if the abnormality is inherited from the mother they will not be affected but the TNDM may occur in the following generation [77].

Genetic testing in neonatal diabetes

At the time of diagnosis of neonatal diabetes it is not known whether diabetes will be transient or permanent. We advise testing for 6q24 abnormalities, KCNJ11, ABCC8, and INS mutations at diagnosis in all diabetes diagnosed before 6 months. Identifying mutations in these genes is important as it will influence treatment. An early diagnosis and very low birth weight make 6q24 most likely. A genetic cause (KCNJ11 or INS)
can be established in approximately 7% of diabetes diagnosed between 6 months and 1 year of age so consideration should be given to testing this age group, especially where autoantibody tests are negative [51].

**Diabetes with extra-pancreatic features**

**Renal cysts and diabetes (RCAD) syndrome: HNF1B-MODY**

**Molecular genetics of RCAD**

HNF1B is a transcription factor with a role in regulating gene expression in a number of tissues. HNF1B forms homo and heterodimers with HNF1A, is present in similar tissues although at different levels, and binds to the same DNA consensus sequence. Despite these similarities HNF1B mutations cause diabetes by a different mechanism and have a distinct phenotype. There is autosomal dominant inheritance, although up to 50% of cases arise from de novo mutations or deletions.

**Phenotype of RCAD**

The predominant phenotype in HNF1B mutations is that of non-diabetic developmental renal disease, most frequently resulting in renal cysts, and diabetes is a less consistent finding. The association of renal cysts with diabetes in HNF1B mutations has led to the designation "Renal Cysts and Diabetes" (RCAD) syndrome [78].

Developmental abnormalities can also be seen in the genital tract, liver, and gut. Uterine and genital tract abnormalities have been described in about 30% of patients [79]. The reason for this is that HNF1B is expressed earlier in embryogenesis and it has a role in regulating gene expression in a number of tissues including the pancreas, kidneys, liver, genital tract, and gut [80]. Birth weight is reduced by around 800 g as a result of reduced insulin secretion in utero [81]. Atrophy of the pancreatic tail can be seen on radiologic scan; associated with loss of pancreatic exocrine function [82] (see Table 28.4). Most cases of HNF1B mutations present to the renal physicians as the varied renal phenotype usually precedes the onset of diabetes. The most common renal phenotype is developmental renal disease, most commonly renal cysts. Renal cysts are present in >75% of patients with HNF1B mutations and are frequently seen on antenatal scanning [51,83]. Three discreet renal histologies have been described: cystic dysplasia, oligomeganephronia, and familial hypoplastic glomerulocystic kidney disease [83]. There are varying degrees of renal dysfunction with approximately 50% of patients ultimately requiring dialysis or transplantation for end-stage renal failure. There is also an association with chromophobe renal cell carcinoma. The severity of the renal disease does not correlate with the genotype [84]. Hypomagnesemia, hyperuricemia, and familial hyperuricemic nephropathy [85–87] have also been described and may reflect a specific defect in renal tubular transport. In marked contrast to HNF1A-MODY, the diabetic people with RCAD have insulin resistance with hypertriglyceridemia. Approximately 50% of cases of RCAD are due to 17q12 gene deletions, and there can be associated learning disability and autism [88].

**Diagnosis of RCAD**

Testing for RCAD should be considered where there is unexplained renal cystic disease (or other renal developmental disease) with or without genito-urinary tract abnormalities or diabetes. As about half the cases of RCAD are spontaneous, lack of a family history should not be a barrier to testing. Testing should not be done in the presence of simple renal cysts and typical type 1 or type 2 diabetes as both conditions are common in the general population. As heterozygous deletions or mutations can cause HNF1B-MODY it is important that copy number analysis, as well as sequencing, is performed when seeking a diagnosis.

**Treatment of RCAD**

Renal disease needs to be managed as any other chronic renal condition, with periodic monitoring of renal function and treatment as appropriate. In addition biennial ultrasounds are recommended to screen for chromophobe renal carcinoma. Patients with diabetes are not sensitive to sulfonylurea therapy and are best treated with insulin.

**Mitochondrial diabetes**

Mitochondria play an essential role in generating chemical energy for cells. Individual cells often contain a variable mixture of mutated and normal mtDNA—termed heteroplasmy. The degree of heteroplasmy varies between tissues and within families and this explains the varied phenotypes seen with the same mitochondrial mutation. Mitochondrial mutations are inherited maternally, as during fertilization the mitochondria are contributed to the offspring by the mother. Maternally inherited diabetes and deafness (MIDD) is the most common mitochondrial diabetes syndrome and will be discussed in the following section. A number of other mitochondrial mutations have been described in families with maternally inherited diabetes (with or without deafness); these are rare (reviewed in [89]).

**Maternally inherited diabetes and deafness (MIDD)**

Maternally inherited diabetes and deafness (MIDD) most commonly results from heteroplasmic G to A substitution of the mitochondrial DNA at nucleotide pair 3243 in one of the two tRNA(Leu) genes [90]. The same mutation that causes MIDD (m.3243A>G) also causes a syndrome of severe neuromuscular disease called MELAS (mitochondrial myopathy, encephalopathy, lactic acidosis, and stroke). Within a family there is usually a dominant phenotype, although occasionally some members develop MIDD whilst others develop MELAS [91]. Most subjects with
MELAS do not develop diabetes, although it has been described. It is not clear why some m.3243A>G carriers develop MIDD whereas others develop MELAS, but it would appear that they are two distinct but overlapping syndromes.

The prevalence of MIDD was 1.3% in a cross-sectional study of the diabetic population in the Netherlands [92]. All children of a mother with an m.3243A>G mutation will inherit the mutation, to some extent. However, only 60% of 3243 mutation carriers will develop diabetes, the remainder seem to have normal glycemic profiles during OGTT in middle age [92]. Spontaneous 3243 mutations have been described, so a lack of family history of diabetes or deafness does not exclude MIDD.

**Phenotype of MIDD**

Diabetes usually develops insidiously in the third to the fifth decade, although the age of presentation is variable, and a fifth of patients present acutely. Pancreatic autoantibodies are usually negative [93]. Over 75% of diabetic m.3243A>G mutation carriers have sensorineural deafness, though this may reflect diagnostic bias (i.e., testing those with diabetes and deafness) [91]. The deafness may present after the diabetes; however, it is usually present in other family members to raise suspicion of MIDD. Hearing loss is more common and more marked in men than women; it is usually loss of high-frequency perception and often requires use of a hearing aid [92] (see Table 28.4).

Other features of MIDD reflect the overlap of this syndrome with MELAS. Neuromuscular signs especially myopathy, high serum lactate or lactate/ pyruvate ratio, nephropathy, cardiac problems with ECG abnormalities, and spontaneous abortions are described [92,93]. Many patients with m.3243A>G mutations have pigmentary retinal dystrophy, although this does not usually affect vision. Patients with MIDD have a high prevalence of renal disease leading to end-stage renal failure, although this can in part be attributed to diabetic nephropathy. The most common histologic finding is of focal segmental glomerular sclerosis, this can predate the diabetes or deafness.

**Diagnosis of MIDD**

The suspicion of MIDD should be raised in any family with diabetes, deafness, or other unusual neurologic features. Diagnosis is made by confirmation of a mitochondrial DNA defect at position 3243. Low heteroplasmic levels in blood can, however, mean that a mitochondrial mutation is missed. If clinical suspicion is high then the molecular diagnosis should be repeated on DNA from urine or oral mucosa, where the heteroplasmacy rates are higher.

**Treatment of MIDD**

Most patients are treated with diet and oral agents at first although there tends to be a rapid progression to insulin, being started at a mean of 2 years after diagnosis. Metformin should probably be avoided because of the overlap with MELAS and risk of metformin precipitating lactic acidosis.

Coenzyme Q10 has been shown to improve respiratory chain function in mitochondria with the m.3243A>G mutation. Anecdotal case reports suggest coenzyme Q10 may prevent hearing loss and delay diabetes in MIDD without side effects, although they mostly just show some improvement in muscle function and lactic acid accumulation on exercise. Maintaining sufficient thiamine intake is thought to be important to optimize mitochondrial function.

**Wolfram syndrome**

Wolfram syndrome is also referred to as DIDMOAD due to the usual occurrence of Diabetes Insipidus, Diabetes Mellitus, Optic Atrophy, and Deafness, although only the presence of diabetes mellitus and bilateral progressive optic atrophy are necessary to make the diagnosis of Wolfram syndrome. The syndrome is rare, with a prevalence of 1.3 per million population in the UK [94].

**Molecular genetics of Wolfram syndrome**

Wolfram syndrome is an autosomal recessive disorder with consanguinity of patients’ parents being common. It is caused by mutation in the WFS1 gene or Wolframin in over 90% of patients [95] and rarely missense mutations in the CISD2 gene [96].

**Clinical features of Wolfram syndrome**

Two case series, comprising about 100 individuals with Wolfram syndrome have given a clear clinical phenotype [94,97]. Diabetes mellitus was the commonest presenting feature (diagnosis median age 6 yrs (range 3 weeks – 16 yrs); 95% were treated with insulin at outset, and C-peptide was low or undetectable. Optic atrophy was diagnosed at a median age of 10 years (range 3 – 30 yrs) and progresses to blindness in most patients. Deafness was present in (46–62%) of patients.

Other features were less common and usually presented later: 50% have diabetes insipidus; 55% renal tract abnormalities (dilated renal outflow tracts and bladder atony); 55% neurologic abnormalities. Males frequently had primary gonadal atrophy and females had menstrual irregularity and delayed menarche but have conceived normal children [94]. There is a high mortality with a median age of death in the UK series of 30 years (range 25 – 49). The most common causes of death were respiratory failure or dysphagia due to brainstem involvement (see Table 28.4).

**Diagnosis and treatment of Wolfram syndrome**

The development of optic atrophy in a young person with insulin-treated diabetes mellitus should raise suspicion of Wolfram syndrome. Other clinical features should be sought and the diagnosis confirmed by molecular genetic analysis, with direct sequencing of the WFS1 gene. There is no treatment to alter outcome of the disease. Diabetes requires insulin treatment with other symptoms being treated as required.

**Using molecular genetics in diabetes care**

In monogenic β-cell diabetes the importance of making a correct diagnosis has seen molecular genetic sequencing move
rapidly from a research tool to clinical care. The use of diagnostic molecular genetics results in new challenges for the clinical diabetes team. Tests are expensive and need to be performed in the patients where monogenic diabetes is most likely. Genetic results need to be correctly reported and interpreted [98]. A molecular genetic diagnosis in a patient needs to be followed by appropriate treatment change and clinical follow-up and also appropriate testing (both biochemical and genetic) for family members.

This change in clinical practice needs to occur in a specialty where there has been very little training in genetics. A source of valuable information for doctors, nurses, and patients can be found at www.diabetesgenes.org.

References

23. Ellard S, Colclough K: Mutations in the genes encoding the transcription factors hepatocyte nuclear factor 1 alpha (HNF1A) and 4 alpha (HNF4A) in maturity-onset diabetes of the young. *Human Mutation* 2006;27(9):854–869.
41 Shih DQ, Dansky HM, Fleisher M, et al.: Genotype/phenotype relationships in HNF-4alpha/MODY1: haplinsufficiency is associated with reduced apolipoprotein (AII), apolipoprotein (CIII), lipoprotein(a), and triglyceride levels. *Diabetes* 2000;49(9):832–837.


CHAPTER 29

Immunopathogenesis of type 1 diabetes in Western society

Alberto Pugliese
Immunology and Microbiology, Division of Diabetes, Endocrinology and Metabolism; Immunogenetics Program, Diabetes Research Institute, University of Miami Miller School of Medicine, Miami, FL, USA

Key points

• Islet autoimmunity appears to be the dominant pathogenic mechanism leading to β-cell loss. Autoimmunity may persist or become reactivated throughout the diabetes clinical spectrum, even decades after diagnosis, for instance when patients receive a pancreas or islet transplant, and favor graft loss.

• Early in life, genetic predisposition impacts the development of thymic tolerance to selected autoantigens. This may represent the earliest pathogenic event; true primary prevention strategies to promote self-tolerance should be instituted in early life, perhaps even during gestation.

• Dysregulation of peripheral immune regulation occurs at multiple levels, promoted by both genetic predisposition and inflammation, in turn amplified by environmental exposures. Many impaired pathways of peripheral immune regulation, as well as inflammation, are being targeted with specific drugs in clinical trials.

• The discovery of several target antigens and the improved understanding of the molecular interactions among antigen-presenting molecules, peptide antigens, and T-cell receptors are facilitating progress towards antigen-specific therapies to selectively delete or regulate autoreactive T cells.

• A critical gap in our understanding of the disease pathogenesis stems from our inability to assess pathology in the pancreas in relation to secretory function. However, growing evidence suggests that loss of β-cell loss and impaired insulin secretion are less severe than previously thought, pointing at additional pathogenic factors and suggesting a wider therapeutic window to reverse diabetes.

Introduction

Type 1 diabetes (T1DM), formerly referred to as insulin-dependent diabetes mellitus (IDDM) and juvenile diabetes, is considered a chronic autoimmune disease. Over time, the disease process results in the virtually complete elimination of pancreatic β cells and lifelong insulin deficiency. In turn, patients become dependent on daily insulin injections to maintain an acceptable level of metabolic control. It is widely accepted that T1DM is a complex, multifactorial disease in which genetic predisposition and environmental exposures promote the triggering of multiple autoimmune responses against β cells.

While a variety of putative environmental factors have been described [1], including infectious agents, dietary and other factors, true etiologic/triggering environmental agents and causal mechanisms remain to be identified. On the other hand, several susceptibility loci have been identified, and many mapped to known genes with predominant function or effects on immune mechanisms, largely supporting a dominant role for an immune-mediated pathogenesis. This is corroborated by the presence of both humoral and cellular autoimmune responses against several islet cell autoantigens near the time of diagnosis. Moreover, in prospective studies, similar responses are observed in nondiabetic first-degree relatives preceding disease development. The most extensively studied autoantigens include insulin/proinsulin, glutamic acid decarboxylase (GAD65), the tyrosine phosphatase-like protein IA-2, the islet-specific glucose-6-phosphatase catalytic subunit-related protein (IGRP), and the cation efflux transporter ZnT8 [2]. Standardized radio-immunoassays measure autoantibodies to insulin, GAD65, IA-2, and ZnT8, which are diagnostic and predictive markers employed as screening tools in natural history studies and prevention trials [3].

The disease commonly develops in children or adolescents, with incidence peaking around 10–14 years of age [4]. Prospective studies in genetically at-risk cohorts demonstrate that islet autoimmunity can be triggered during early childhood [5]. However, T1DM is also diagnosed in adults [4]. Some adult patients with milder presentation are diagnosed with latent onset diabetes of the adult (LADA); it has been proposed that
LADA may fundamentally be a clinical expression of T1DM characterized by later triggering and/or slower progression [6], perhaps occurring in individuals with attenuated genetic susceptibility [7]. The disease incidence and prevalence vary in children from different countries and populations within an approximate 350-fold spread [4]. The lowest incidence rates (0.1–1/100,000) are reported in Asian populations (China, Japan) and some South American countries. The highest incidence rates, ranging from 5 to 60/100,000, are reported in Caucasian populations of the Western world [8]. Northern European countries such as Finland and Sweden have very high incidence and prevalence, but a similarly high incidence is reported in Sardinia, an island in the Mediterranean Sea [9]. During the past few decades the worldwide disease incidence has been rising 2–5% annually, albeit not uniformly in various countries and throughout time periods [8]. Moreover, many studies indicate that T1DM is becoming more common in younger children (<5 years old) [4,8,10]. The prevalence of T1DM in the United States, where the population comprises multiple racial and ethnic groups, is approximately 1/300 by age 18; data from the SEARCH study indicate that incidence and prevalence rates are higher (roughly by 30–50%, depending on ethnic and age groups compared) in non-Hispanic whites (>20/100,000 in youth less than 20 years old) compared to Hispanic, African Americans, and American Indians [4].

Overall, epidemiologic data show that T1DM is more prevalent in the Western world, especially affecting, albeit not exclusively, Caucasians of Northern European descent. This chapter focuses on reviewing recent and major findings about the disease pathogenesis that are largely originated by studies in patients from the Western society.

**Genetic studies provide clues about key disease mechanisms**

Genetic predisposition is an important component of the multifactorial pathogenesis of T1DM. Although often diagnosed in individuals with no known family history of T1DM, the disease is about 15 times more common in siblings of a patient than in the general population. Siblings have an average risk of 6%, although individual risk varies significantly in relation to the extent of sharing predisposing genes with the proband, which allele variants are shared, and other factors. The risk to the offspring of affected mothers and fathers is about 2–3% and 6–7%, respectively [4]. Among twins, the observed disease concordance rates are approximately 8–10% in dizygotic twins and, with extended follow-up, more than 60% in monozygotic twins [11].

Extensive genetic studies have been conducted during the past decades in thousands of families with affected sibling pairs, family trios, patients, and unrelated controls, including genome-wide association studies. The largest coordinated efforts were led by the Type 1 Diabetes Genetic Consortium (www.T1DGC.org), an international collaboration [12]. A single, major susceptibility locus and a multiplicity of other loci (upwards of 50; see http://www.t1dbase.org), individually conferring a much smaller risk, have been identified [12]; indeed, only four of these genes (HLA Class II, insulin, PTPN22, and IL2RA) are associated with odds ratios greater than 1.5 [12]. Several, but not all, of the genetic loci associated with T1DM have been confirmed in multiple data sets and populations [13]. Many if not most of the susceptibility loci mapped to known genes encode for proteins involved in immune function; a few other genes are expressed in pancreatic islets and putative risk genes remain to be mapped at some recently associated loci [12]. Besides inherited alleles, additional mechanisms may regulate gene expression and function, including epigenetic regulation, typically mediated by changes in DNA methylation [14]; in addition to contributing to susceptibility, epigenetic regulation may represent a mechanism by which gene–environment interactions can impact disease risk and could help explaining the significant increase in disease incidence reported in many populations [15]. In the next paragraphs we will discuss putative disease mechanisms associated with the best studied risk genes. Figure 29.1 illustrates these mechanisms and compares the effects of predisposing and protective gene variants at the HLA-DQ, insulin, PTPN22, CTLA-4, IL2RA, and IHI61 loci.

**HLA-encoded genetic predisposition points at adaptive autoimmune responses as the dominant effector mechanism leading to β-cell destruction.** As noted, the identification of several susceptibility genes involved in key immunologic functions supports a key role for autoimmunity in the disease pathogenesis. Many of these genes control critical functions of the adaptive immune system, in particular, T-cell responses and their regulation. Experimental evidence points at T cells as the main effectors of β-cell destruction. Accordingly, the primary susceptibility locus maps to the human leukocyte antigen (HLA) complex, specifically to HLA class II and class I antigen-presenting molecules; these function as restricting elements for CD4 and CD8 T-cell responses, respectively. HLA-encoded susceptibility confers up to 50–60% of the overall genetic risk from inherited alleles [16].

Within the HLA complex, HLA class II loci have the strongest association with disease risk; in particular, the HLA-DR3, DQA1*0501-DQB1*0201 (DQ2) and HLA-DR4, DQA1*0301-DQB1*0302 (DQ8) genotype is the strongest susceptibility factor with an odds ratio >40 [12]. This genotype is typically carried by 30–50% of patients, compared to about 2% of the general population; most patients carry at least one of these two high-risk haplotypes. The higher risk conferred by the heterozygous genotype is explained by the formation of a trans-complementing HLA-DQ heterodimer, which can only be formed when the two haplotypes are inherited together, that is capable of presenting islet autoantigens and promote diabetogenic responses [17]. On the other hand, HLA-DR4, DQA1*0301-DQB1*0301 (DQ7) haplotypes are considered neutral. This observation links diabetes risk with the DQB1*0302 chain. In contrast, HLA-DR2 (DRB1*1501)
Figure 29.1 Best characterized genetic mechanisms that influence islet autoimmunity. The figure illustrates and contrasts the molecular mechanisms underlying genetic susceptibility and genetic protection for well-characterized loci (HLA, insulin, PTPN22, CTLA-4, IL2RA, and IFIH1). The interaction between a CD4 T cell and an antigen-presenting cell (APC) is used to depict these mechanisms.

**HLA class I/II:** presentation of autoantigen peptides to CD8/CD4 T cells in thymus and periphery is influenced by the HLA variant;

**Insulin:** predisposing variants of the insulin gene VNTR are linked to reduced insulin expression in thymus, and in turn to impaired negative selection of insulin-specific T cells;

**PTPN22 (Lyp):** the predisposing variant identified at this locus is a gain of function mutant, resulting in increased inhibition and impaired TCR signaling, which may affect multiple processes including thymic selection of T cells;

**CTLA-4:** predisposing variants associated with reduced levels of sCTLA-4, may impair control over T-cell activation through effects on regulatory T cells;

**IL2RA:** predisposing variants associated with low levels of sIL2RA, reduced STAT5a responses in antigen-experienced CD4 T cells, and impaired regulatory T-cell function with decreased expression of the FOXP3 transcription factor (not illustrated);

**IFIH1:** predisposition is associated with increased expression of IFIH1, IFN, and NFκB responses, cytokines, and an overall enhanced inflammatory response from APCs, following recognition of viral RNAs; this locus links susceptibility to viral infections.

haplotypes carrying DQA1*0102-DQB1*0602 are negatively associated with T1DM in several populations, with less than 1% of T1DM patients carrying this protective haplotype; reportedly, this haplotype confers strong protection from T1DM development also in first-degree relatives with autoantibodies [18].

Several HLA-DR4 subtypes are defined by polymorphisms of the DRB1 chain, for example DRB1*0401, DRB1*0404, and so on, and they show variable degrees of association with T1DM risk, even when in linkage with DQA1*0301-DQB1*0302; this suggests that the HLA-DRB1 molecule also plays an important role in modulating immune responsiveness to islet antigens [16]. While there is significant linkage disequilibrium across the HLA complex, selected HLA class I alleles are also associated with T1DM, independently of class II genes, especially HLA-A2, HLA-A24, and HLA-B39 [19]. Many studies suggest that additional loci within the HLA complex may modulate susceptibility. The largest and most systematic evaluation was conducted by the T1DGC, which reported data for more than 3000 single nucleotide polymorphisms, 66 microsatellites, and class I and class II DNA-based allele typing data in 2300 families with affected sibling pairs; these studies identified several associations involving many genes within the HLA complex, suggesting that HLA-encoded susceptibility derives from multilocus effects [20].
Genetic predisposition impacts central immune tolerance to islet cell antigens. Antigen presentation by HLA molecules is not only a key to the activation of adaptive immune responses, but also to the generation of regulatory T cells and the stimulation of other regulatory mechanisms. Antigen presentation is also required for T-cell selection processes occurring in the thymus during the maturation of the immune system to establish immunologic self-tolerance. Genetic predisposition can impair central (thymic) tolerance to selected autoantigens, and this may well represent the earliest event in T1DM pathogenesis.

Such mechanisms became apparent following the discovery that peripheral or tissue-restricted antigens (TRAs) from multiple organs are "ectopically" produced in the thymus to support the development of immunologic self-tolerance in early life [21]. TRA expression in the thymus is primarily mediated by medullary thymic epithelial cells (mTECs), which transcribe many TRA genes under the control of the AIRE autoimmune regulator transcription factor [22]. Thymic CD11c+ dendritic cells (DCs) also express insulin and other TRAs [23], but TRA expression levels are lower in DCs compared to mTECs [24]. DCs can also uptake antigens produced by mTECs [25], a function that is also supported by AIRE [22].

Insulin itself is a prototypical TRA, its synthesis being virtually restricted to pancreatic β cells. Thymic insulin production is critical for establishing self-tolerance to β cells; simply abolishing insulin expression in the thymus leads to the rapid onset of autoimmune diabetes even in mice lacking a diabetogenic genetic background [26]. In humans, thymic insulin expression is modulated by allelic variation and epigenetic effects at the insulin gene locus; this effect is largely mediated by a polymorphic variable nucleotide tandem repeat (VNTR) sequence. In the thymus, protective VNTR alleles are associated with 2.5- to 3-fold higher insulin gene transcription levels compared to predisposing alleles [27,28]; polymorphisms of the AIRE and MafA transcription factors also modulate insulin gene transcription in the thymus [29,30]. As with other self-molecules, thymic insulin levels impact negative selection of autoreactive T cells and/or positive selection of regulatory T cells; lower levels typically lead to less efficient selection processes, promoting the maturation of a T-cell repertoire that is enriched in insulin autoreactive T cells [31].

Immunologic tolerance to other islet autoantigens is likely shaped by thymic expression as well: splice variants of the IA-2 and IGRP islet autoantigens have expression patterns that differ in human pancreas and thymus, suggesting that self-tolerance to antigenic epitopes expressed only in the pancreas may not be fully achievable [32,33]. Thus, allelic variation, alternative splicing, and epigenetic regulation of gene expression can affect quantity and quality of TRA expression in the thymus, in turn impacting thymic selection processes and favoring the maturation of a T-cell repertoire that is prone to autoimmune responses to a variety of islet autoantigens.

Escape from negative selection may also be favored, albeit not in an antigen-specific fashion, by polymorphisms of the PTPN22 gene. This encodes for Lyp, an intracellular, lymphocyte-specific tyrosine phosphatase, which functions as a negative regulator of TCR signaling. PTPN22 predisposition may derive from more potent suppression of TCR signaling during thymic development, resulting in impaired negative selection [34]. Furthermore, recent investigations of patients’ autoreactive CD4+ T cells highlight abnormal features of the immunologic synapses that may also favor escape from negative selection and promote effector responses in the target organ, where larger amounts of antigen are present [35].

Overall, there is strong evidence that the earliest disease mechanisms and primary genetic predisposition effects are active in early life, through modulation of thymic selection processes, with the stronger effects likely being specific for selected islet cell molecules, primarily under the genetic control of HLA and insulin genes. Therefore, true primary prevention strategies should aim at improving the efficiency of thymic selection processes and should be attempted very early, perhaps in utero. For example, drugs are being sought to enhance thymic insulin expression [36], which could be given to genetically at-risk children if safe and effective.

Functional properties of HLA class II and class I molecules presenting diabetogenic peptides. As noted, HLA class I and class II molecules mediate TRA presentation to developing lymphocytes in the thymus. In patients with T1DM, these will commonly be the predisposing variants described earlier. Molecular studies reveal that certain amino acid residues impact the peptide binding properties of these molecules and in turn presentation of islet antigen peptides to autoreactive T cells; for example, DQB1*0302 carries a serine at position 57, where most nonpredisposing HLA-DQ molecules carry an aspartic acid residue [37]. There are remarkable sequence (including position 57) and structural similarities between the human HLA-DQ8 molecule and its homologous I-Ag7 in the MHC (major histocompatibility complex) of the NOD (nonobese diabetic) mouse, a model of autoimmune diabetes. There are also similarities in the binding of insulin and other peptides to these two molecules [37]. Recent landmark studies in the NOD mouse show that I-Ag7 binds a diabetes-associated insulin peptide (B9-23) in multiple registers: in the thymus, a register binding is used that results in poor negative selection of these autoreactive T cells [38], while presentation to autoreactive T cells in the periphery may use a different binding register that will trigger autoimmune responses [38]. Of note, insulin is the only islet autoantigen shown to be critical for diabetes development in the NOD mouse [39]; therefore, these studies reveal that the molecular basis of a key disease mechanism is explained by the binding features of the predisposing MHC molecule. Likewise, in humans, the molecular interactions of a pre-proinsulin peptide with the HLA-A2 molecule are weak, resulting in low TCR-peptide-HLA binding affinities. These studies illustrate how the molecular features of predisposing HLA molecules can result in poor presentation of a pre-proinsulin peptide and in enhanced probability that
autoreactive CD8 T cells may escape thymic selection and later become autoreactive [40]. The discovery of target antigens and the definition of key molecular MHC-peptide interactions are critical for developing antigen-specific therapies to selectively delete or regulate autoreactive T cells [41].

Genetic predisposition also impairs peripheral tolerance to islet autoantigens. An immunologically relevant expression of insulin and other peripheral antigens is not limited to the thymus, but it is now well described in peripheral lymphoid tissues as well. Human DCs express insulin and several other TRAs [23,42]; studies in mice demonstrate that insulin expression by AIRE-expressing DCs maintains peripheral tolerance to this critical diabetes autoantigen [43]. Several other cell types have been linked to TRA expression in the periphery [24,44,45]; for example, lymph node stromal cells express many TRAs under AIRE transcriptional control and mediate deletion of autoreactive T cells specific for a model autoantigen in transgenic mice; however, these cells do not express insulin [24]. Thus, the immune system has mechanisms ensuring redundancy and complementarity of peripheral TRA expression, to help maintain peripheral tolerance after the inversion of the thymus. It is plausible that the same predisposing genes (HLA, insulin, PTPN22), alternative splicing, and epigenetic mechanisms that impact thymic selection may also impart similar influences on TRA expression in the periphery. In the NOD mouse, insulin gene expression in the pancreatic lymph node (PLN) is lost in relation to proinflammatory changes developing as mice age and progress towards diabetes [46]. These studies link reduced insulin expression to another transcription factor, Deaf1, which is alternatively spliced into a less active form in the PLN of NOD mice; similar findings have been replicated in the pancreatic lymph nodes of patients with T1DM [47]. These observations link inflammation to impaired peripheral tolerance, via effects on TRA expression of a critical islet autoantigen. Then, perhaps, anti-inflammatory therapies may have a role in disease prevention and/or treatment if they can help preserve peripheral tolerance mechanisms mediated by the presentation of self-antigens.

Multiple risk loci impair peripheral immune regulation. Non-antigen-specific abnormalities of peripheral tolerance are also involved in the pathogenesis of T1DM [48]. Multiple defects impair the function of patients’ regulatory T cells and patients’ effector T cells are resistant to suppression by regulatory T cells [49]. There is also an imbalance between Th17 immunity and regulatory T cells both in peripheral blood and the PLN of T1DM patients [50]. Several susceptibility loci promote enhanced immune reactivity and less effective control over T-cell selection, activation, and perhaps differentiation into memory and regulatory phenotypes [12]. The effects of these loci are not antigen- or disease-specific; indeed, these loci confer increased risk for several autoimmune disorders. The effects of a selected few, PTPN22, CTLA-4, and IL2RA are summarized here and illustrated in Figure 29.1. Through its effects on TCR signaling, the PTPN22 locus affects T-cell function in the periphery, and reports indicate that it also impacts the function of regulatory T cells and B lymphocytes [51,52]. The CTLA-4 (Cytotoxic T lymphocytes- antigen-4; CD152) gene encodes for a protein, expressed by activated CD4 and CD8 T cells, which downregulates T-cell responses. CTLA-4 polymorphisms linked to T1DM are associated with reduced levels of a soluble form of CTLA-4 with a negative impact on regulatory T-cell function [53]. Of note, a recent trial showed that CTLA-4-Ig therapy mitigated loss of insulin secretion in new onset patients, albeit the effect was limited in time [54]. Polymorphic variants of the IL2RA gene (coding for the α chain of the Interleukin-2 receptor, IL-2Rα, or CD25) modulate T1DM risk through effects on several functions, including reduced levels of soluble IL-2 receptor [55], reduced STAT5a signaling in antigen-experienced CD4 T cells, and impaired regulatory T-cell function with decreased expression of the FOXP3 transcription factor [56]. Studies from both NOD mice [57] and other autoimmune diseases indicate that impaired IL-2 signaling compromises T regulatory function. Hence, there is interest in using low dose IL-2 to ameliorate autoimmunity by selectively stimulating regulatory T cells [58]. Overall, imperfect regulation of peripheral tolerance mechanisms, largely but not exclusively through effects on regulatory T cells, plays an important role in T1DM pathogenesis.

Genes and environment: a moving target?

Studies from multiple populations have shown shifts of the typical HLA class II gene associations during the last few decades. Indeed, fewer patients carry the high risk heterozygous HLA-DR/DQ genotype, except among younger children [59–63], in whom the disease is becoming more common [4]. In contrast, more patients now carry moderate risk HLA types and genotypes. Such shifts in HLA associations may be explained by stronger environmental pressures that enhance HLA-mediated genetic predisposition and/or broaden the spectrum of diabetogenic gene–environment interactions. Viruses, and especially enteroviruses, rank at the top of the list of environmental factors that have been linked to T1DM [1].

Several lines of evidence support a role for viruses [64]:
1. diabetes incidence correlates with the seasonality of enterovirus infections;
2. reports of T1DM epidemic bouts;
3. associations of T1DM with immune responses to enterovirus and detection ofenteroviral RNA, which correlate with islet autoimmunity and β-cell function [65];
4. reports demonstrate higher T1DM risk in the offspring of mothers with enteroviral infections [66] and, related to this observation, enteroviruses may infect the thymus and alter thymic selection processes in turn favoring autoimmunity [67];
5. the identification of a susceptibility gene controlling innate responses to enterovirus [68]: the IFIH1 gene encodes for the interferon induced with helicase C domain protein 1, which
recognizes enteroviral double-stranded RNA (dsRNA) in infected cells and promotes interferon and NFκB responses, followed by the production of inflammatory cytokines. Experimental data suggest that viral infection may lead to sustained interferon responses and upregulation of HLA class I expression in the presence of predisposing IFIH1 alleles, which would in turn increase the potential to expose self-antigens and trigger autoreactive responses. In contrast, T1DM-protective IFIH1 variants may be associated with reduced responses and decreased likelihood to trigger autoimmunity (Figure 29.1);

6 growing evidence that enterovirus can infect β cells in T1DM patients [69,70]; studies in large cohorts of deceased donors show that the presence of viral antigen is detected much more frequently in the pancreas of T1DM patients compared to patients with T2DM and nondiabetic subjects [70]. Further, enteroviruses can infect and damage β cells, induce expression of HLA class I antigens and alpha-interferon [69], which could amplify inflammation and provide a link to the triggering of islet autoimmunity. Enterovirus infections may also induce functional changes and influence β-cell replication [65,71]. Researchers are testing the hypothesis that multiple, acute viral infections and/or chronic viral infections may repeatedly trigger and sustain autoimmunity over time. Future studies could possibly link one or more serotypes to T1DM, with obvious implications for improved prediction, prevention, and treatment against the triggering and progression of islet autoimmunity.

Multiple T-cell types and autoreactive T-cell specificities mediate T1DM

The inflammatory lesion known as insulitis is considered the pathologic hallmark of T1DM, typically seen during the prediabetic stage or around the time of onset: it consists of the infiltration of the pancreatic islets by lymphocytes and other inflammatory cells. Of great significance is the observation that β cells of NOD mice contain granules enriched in a diabetogenic insulin peptide (B9-23), which is up-taken by intra-islet DCs and presented to CD4 T cells in the PLN [72,73]. This finding calls attention to the β cell itself as the source of a peptide antigen which, at least in the NOD mouse, is the main driver of islet autoimmunity [39]. Once activated in the PLN, autoreactive CD8 T cells recognize MHC class I restricted islet autoantigens directly on the β-cell surface and mediate their cytotoxic effects through multiple effector molecules, including perforin, Fas ligand, and Th1 cytokines, such as IFN-γ and TNF-α. CD4 T cells are linked with the triggering of the autoimmune process and support activation and recruitment of cytotoxic CD8 T cells into the islets. However, CD4 T cells also stimulate β-cell killing via the secretion of cytokines and via signaling through death receptors (Fas, TNF-α) [73]. While T1DM is traditionally considered to be mediated by Th1 type T cells, growing evidence implicates Th17 T cells, which have been shown to have deleterious effects on β cells [74].

In both patients and animal models, studies have identified multiple CD4 and CD8 T cells reacting against a variety of islet autoantigens [2], including multiple epitopes from the autoantigens insulin/proinsulin, GAD65, IA-2, IGRP, and ZnT8. In patients, most studies are limited to peripheral blood samples; however, progress has been made in validating that these cells are linked to disease pathology in the target organ: (1) insulin autoreactive T cells were found in the PLN of T1DM patients [75]; (2) in pancreas transplant recipients, proinsulin/insulin and GAD-specific CD4 T cells were detected in both the circulation and the pancreas transplant lymph node in biopsies in association with recurrent diabetes, as demonstrated by insulitis in the transplanted pancreas [76]. Recent studies also provide functional evidence for the pathogenic role of human autoreactive T cells: (1) GAD-reactive CD4 T cells from the above pancreas transplant patients with recurrent diabetes killed human islets, in vivo, when co-transplanted under the kidney capsule of immunodeficient mice [76]; (2) proinsulin-specific CD8 T cells from the peripheral blood of T1DM patients kill β cells in vitro [77]; (3) autoreactive T cells were directly demonstrated in the insulitis lesion, in the pancreas of an organ donor with T1DM, using MHC tetramers [78].

Autoimmunity persists in many patients even many years after diagnosis

Studies in patients with long disease duration suggest that β-cell function may persist long after diagnosis [79,80] in a significant proportion of patients. There is also evidence suggesting the co-existence of some low level of regeneration with chronic autoimmunity [81], as islet autoimmunity may also persist or perhaps be reactivated years after diagnosis. Indeed, significant proportions (30–40%) of islet or pancreas transplant recipients express one or more autoantibodies when evaluated prior to transplantation [82]. Moreover, persisting autoreactive T-cell responses detected prior to transplantation correlate with islet graft failure on follow-up [83]. These cells may be expanded following transplantation, since the lymphopenia induced by chronic immunosuppression may promote homeostatic proliferation of memory T cells, which in patients with an autoimmune disease will include autoreactive T cells [84]. The persistence and/or reactivation of islet autoimmune responses is an obstacle to curing diabetes through transplantation, and autoimmunity may not be completely controlled by the immunosuppression used to prevent transplant rejection [76]. Thus, islet autoimmunity may be active for many years, throughout the diabetes clinical spectrum (Figure 29.2), and not just limited to the time prior to and around diagnosis.

Nonautoimmune factors

Additional factors may play a role in the disease pathogenesis, and may promote β-cell dysfunction and progression of islet
autoimmunity. Insulin resistance affects about 20% of young patients with T1DM, especially during puberty. Importantly, insulin resistance precedes T1DM diagnosis and accelerates the progression of islet autoimmunity [85–88]. Obesity and T2DM have become more common in children [89]; children with T1DM are often overweight [89], suggesting that certain pathogenic mechanisms may be shared between T1DM and T2DM [90]. Significant changes in metabolite profiles (amino acids, lipids, fatty acids) precede islet autoimmunity in genetically at-risk children, which could be linked to inflammation and possibly apoptosis occurring in the pancreas [91,92], which has been linked to the PTPN2, a susceptibility gene expressed in islet cells that control apoptosis [93]. Insulin resistance and metabolic changes may be also be related to inflammation. In the SEARCH study, young patients with T1DM had elevated serum levels of IL-6, fibrinogen, and C-reactive protein [94]; stimulation of patients’ monocytes via toll-like receptors is followed by exaggerated IL-6 responses [95]. Of note, IL-6 has been linked to increased Th17 T-cell responses and impaired regulatory T-cell function [96]. The serum of T1DM patients is characterized by a proinflammatory signature that includes IL-1 and several chemokines [97], many of which were also expressed in the pancreas of T1DM patients [98]. Proinflammatory cytokines have deleterious effects on β-cell function and survival [99], and pancreatic islets from patients with T1DM reportedly express markers of endoplasmic reticulum stress [100]. Inflammation is often associated with obesity, and can be modulated by genetic and environmental factors. In addition to a possible role of enteroviral infections in promoting inflammation in the gut [101], there is growing evidence that dietary habits can influence intestinal permeability and the composition of the gut flora [102]; this in turn can influence production of cytokines, chemokines [98], innate immune responses [95], and the function of gut-associated lymphocytes, including regulatory lymphocytes.

Overall, growing evidence points at inflammation and metabolic changes as significant factors in T1DM pathogenesis. These could even precede the triggering of islet autoimmunity, and may promote chronic immune dysregulation through effects on both innate and adaptive immune responses. Of note, some of the genes linked to T1DM risk include genes that influence the innate immune system [12]. Further genetic studies

The T1DM Spectrum

<table>
<thead>
<tr>
<th>Genetic Susceptibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Environmental Factors</td>
</tr>
<tr>
<td>Prediabetes</td>
</tr>
<tr>
<td>Onset</td>
</tr>
<tr>
<td>Complications</td>
</tr>
</tbody>
</table>

**Figure 29.2** The T1DM clinical spectrum. Illustration of the various clinical stages of the diabetes clinical spectrum and of the biologic phenomena that are the subject of intense research.
may provide a genetic underpinning for these and additional disease pathways, considering that not all of the T1DM risk loci map to immune function genes, that some susceptibility genes are expressed in islet cells, and that additional susceptibility genes remain to be mapped [12].

The relationship between pathogenic factors and β-cell destruction remains poorly understood

Longitudinal studies of newborns and children at genetic risk [5,103,104] and follow-up of at-risk first-degree relatives in natural history studies link the triggering of autoimmunity with the appearance of autoantibodies to one or more islet autoantigens; individuals with multiple autoantibodies have higher risk of diabetes progression [3]. Data from the Diabetes Prevention Trial-Type 1 document a progressive impairment of insulin secretion and glucose metabolism as subjects progressed towards diagnosis [105]. Recent clinical studies have led to increased awareness that stimulated C-peptide responses are only partially reduced in many newly diagnosed patients, to an extent as a function of age [106,107]. A two-year follow-up of new onset patients shows greater C-peptide loss during the first year post diagnosis; importantly, not every patient experienced further loss during this period [108]. At least low levels of C-peptide production may persist for several years in a significant proportion of patients, even decades after diagnosis [79,80]. In parallel, it has been proposed that physical β-cell loss at diagnosis may be also age dependent and, importantly, less severe than previously estimated; a recent meta-analysis estimates the average β-cell loss by age 20 at approximately 40% [109]. The effect of age of onset on residual β-cell mass may reflect age-related differences in the number of β cells, but also the influences on β-cell mass and diabetes progression of body mass index and insulin resistance, which increase as children go through adolescence and puberty. Related to the above, the severity of insulitis from pathology specimens also appears to be a function of age, besides disease duration [110]. Insulitis in the human pancreas is generally not as severe and extensive as observed in NOD mice; even among younger patients with short disease duration, who typically have the most severe insulitis and β-cell loss, only about 30% of insulin-positive islets had a lymphocytic infiltrate [110]. Moreover, recent studies described heterogeneous β-cell loss patterns in pancreas pathology specimens from patients diagnosed in childhood with variable disease duration [111]. While it is not possible to correlate insulin secretion with an assessment of β-cell mass in patients with new onset diabetes, studies in pancreas transplant patients with recurrent diabetes, in whom a pancreas transplant biopsy was obtained, often showed moderate β-cell loss and insulitis despite a frank diabetic state [76]. Collectively, these observations raise the provocative question of whether additional co-factors may be impairing β-cell function at the time of diagnosis.

Conclusions

Both genetic and immunologic studies show that autoimmunity plays a major role in T1DM pathogenesis. The molecular characterization of several autoantigens and disease-associated gene variants has led to the identification of several disease pathways and functional abnormalities that contribute to the immune-mediated component of T1DM pathogenesis, many of which are being targeted in clinical trials. Islet autoimmunity appears to be the dominant effector mechanism. Over time, chronic autoimmunity will lead to the virtually complete loss of pancreatic β cells in many but not all patients, given that C-peptide production is detected in a significant proportion of patients, even decades after diabetes onset. There is growing evidence that additional factors contribute to disease pathogenesis and progression; moreover, re-examination of both literature data and emerging findings suggest the hypothesis that a degree of functional impairment may contribute to diabetes symptoms at diagnosis, at least in some patients. If so, additional pathogenic mechanisms and in turn therapeutic targets may exist at diagnosis. Ultimately, a critical gap in our understanding of the disease pathogenesis stems from our inability to assess pathology in the pancreas in relation to secretory function. This prevents us from fully understanding the relationships among insulin secretion, β-cell mass, disease activity, and relative contributions of various factors to disease pathogenesis and clinical symptoms. Longitudinal assessments of T-cell responses during the prediabetic phase are still quite limited; with recent technical improvements future studies should illustrate the dynamics of T-cell responses, antigen specificity, and functional properties of autoreactive T cells in relation to disease progression. Correlations with biomarkers of β-cell destruction have not yet been possible, as another unmet need is a validated biomarker; the levels of circulating, demethylated insulin gene DNA levels could represent such a biomarker if validated in larger studies [112]. There is also hope that further progress will lead to advanced imaging modalities to noninvasively assess β-cell mass and insulitis [113]. This need is even more critical now that improved knowledge of the disease natural history shows relative preservation of C-peptide secretion at diagnosis and later follow-up in many patients, suggesting that the therapeutic window for intervention may extend beyond the first few months from clinical diagnosis.

References


7 Grant SF, Hakonarson H, Schwartz S: Can the genetics of type 1 and type 2 diabetes shed light on the genetics of latent autoimmune diabetes in adults? *Endocrine Reviews* 2010;31(2):183–193.


CHAPTER 30
Molecular genetics of type 1 diabetes

M. Ann Kelly1,2, Simon D. Rees1,2, Anthony H. Barnett1,2 and Stephen C. Bain3
1Clinical and Experimental Medicine, College of Medicine and Dentistry, University of Birmingham, Edgbaston, Birmingham, UK
2Diabetes Centre, Heart of England NHS Foundation Trust, Birmingham, UK
3Institute of Life Science, College of Medicine, Swansea University, Swansea, UK

Key points
- Type 1 diabetes (T1DM) is a complex disease, with both genetic and environmental factors contributing to overall disease risk.
- To date, over 50 genetic loci have been implicated in T1DM risk by a mixture of candidate gene analysis and genome-wide association studies. In combination, however, these account for only 70% of the heritability of the disease. The remaining 30% has yet to be explained.
- The largest genetic effect is conferred by the HLA gene region on chromosome 6p21. This alone accounts for approximately 40–50% of disease heritability. Disease risk cannot be attributed to a single locus within this region, however. Genetic and functional studies suggest that the majority of HLA-mediated risk is conferred by the HLA-DRB1, -DQA1 and -DQB1 loci, although alleles at the -DPB1, HLA-A, HLA-B, and other loci also modulate disease susceptibility.
- The greatest risk of T1DM is conferred by the DRB1*0301.DQA1*0501.DQB1*0201/DQB1*0302 genotype (termed DR3.DQ2/DR4.DQ8), which increases the odds of developing the disease approximately 16-fold. In contrast, strong dominant protection is conferred by DRB1*1501.DQA1*0102.DQB1*0602 (DR15.DQ6). A range of other HLA haplotypes have more moderate effects on disease risk.
- Sequence variants in the insulin gene (INS) and PTPN22 gene increase the risk of T1DM approximately twofold.
- The remaining susceptibility loci have more modest effects on disease risk (modifying risk by 5–30%) and, in the majority of cases, the causal variants remain to be identified.
- Current understanding of the genetic basis of T1DM has provided valuable insights into the mechanisms underlying the loss of T-cell tolerance to β-cell autoantigens and identified promising therapeutic targets for disease prevention.
- Further research is necessary to fully define the genetic background of T1DM. Future association studies of sequence variants should be complemented by functional genomics and mechanistic studies to elucidate the molecular pathways underlying the pathogenesis of T1DM and identify putative disease loci along these pathways.

Introduction
Type 1 diabetes (T1DM) is a chronic autoimmune disease in which the β cells of the islets of Langerhans are selectively destroyed, resulting in insulin deficiency and hyperglycemia. The disease develops in genetically susceptible individuals, most likely as a result of an environmental trigger. T1DM has an uneven geographical distribution; disease prevalence is highest in populations of white European origin and lowest in those of East Asian descent [1]. A marked gradient in disease risk also exists in Europe, with higher prevalence of T1DM in northern countries, particularly Finland, compared with areas around the Mediterranean. This pattern could be attributed to genetic differences between the populations or to the presence/absence of environmental triggers.

The sibling relative risk (λs) measures the probability of a disease developing in a sibling of a proband compared with the risk in the general population. This index of familial clustering is often used as a measure of genetic effect, although it could also reflect the impact of shared environmental exposures. The λs for T1DM is 15 (sibling risk 6%/population risk 0.4%), one of the highest values observed among common complex diseases, suggesting a substantial inherited component to T1DM [2]. This is further supported by the high concordance rate observed among monozygotic (MZ) twins (>50%), while concordance in dizygotic (DZ) twins is similar to that seen for other siblings (6–10%) [3]. A long-term follow-up study of initially discordant MZ twins showed that 65% of the nonproband co-twins developed T1DM by the age of 60 years, while 78% developed persistent islet autoimmunity, suggesting that genetic susceptibility persists for life [4]. Concordance does not reach unity, however, and there is significant divergence between twins in the time taken to develop disease. This implies that environmental factors are also important. Epidemiologic evidence of a
marked increase in T1DM incidence in many areas of the world over the past 20–30 years [5,6] also supports a significant contribution from environmental risk determinants, as do studies of migrant populations, which suggest that individuals from a low-prevalence country can increase their risk of developing T1DM if they move to an area of higher prevalence [7]. This increase in risk may be attenuated by the individual's genetic background, however, such that they do not acquire as high a disease risk as the indigenous population of their adopted country [8,9]. Conversely, children from high-risk populations who are born and raised in areas with lower T1DM prevalence (such as those of Sardinian ancestry raised in Italy) retain their higher risk of disease compared with the host population, underlining the importance of genetic makeup to T1DM risk [10].

Type 1 diabetes is a polygenic disease, in which a large number of susceptibility loci contribute to overall disease risk. The susceptibility genes have low penetrance and, as a result, not all individuals judged to be “genetically at-risk” will develop the disease. Furthermore, diabetes can occur in the absence of known high-risk markers. The lack of a simple relationship between genotype and phenotype makes it difficult to identify disease genes at a population level and studies rely on the detection of statistical associations that are unlikely to have occurred by chance. To date over 50 genetic loci have been shown to be reproducibly associated with T1DM [11, http://t1dbase.org]. In most of these regions, however, fine-mapping is required to define the specific gene(s) involved and identify the causal variants. This chapter reviews our current understanding of the genetic basis of the disease and discusses the implications for disease prediction and the identification of novel targets for intervention in the development and progression of islet autoimmunity.

The HLA complex

The genes of the human leukocyte antigen (HLA) complex were first identified as important determinants of T1DM risk in the 1970s [12]. Subsequent family studies comparing disease concordance rates between HLA-identical siblings and monozygotic twins suggested that the HLA genes are the major genetic contributor to T1DM, accounting for about half of the familial aggregation of the disease [13]. Recent studies concur with this estimate and no other locus with such a substantial influence on disease risk has been identified [2,11].

The HLA complex (also known as the human major histocompatibility complex, MHC) maps to a 3.6Mb region on chromosome 6p21.31 and consists of more than 200 identified genes, over half of which are known to be expressed. The HLA class I genes are located at the telomeric end of the complex, with the class II genes at the centromeric end. The 700 kb sequence between these gene clusters is commonly referred to

![Figure 30.1](image-url) - A simplified map of the human major histocompatibility complex (MHC) on chromosome 6p21.31. The genes reported to be reproducibly associated with T1DM in several independent datasets are shown in dark gray.
as the HLA “class III” region, although it contains no classical HLA genes (Figure 30.1). The HLA complex plays a crucial role in the immune response, in particular the genes encoding the classical class I (HLA-A, -B, and -C) and class II (HLA-DR, -DQ, and -DP) molecules. The HLA-A, -B, and -C genes each encode a peptide chain that combines with β2 microglobulin to form the HLA class I molecules, expressed on the surface of all nucleated cells (Figure 30.2(a)). These molecules bind to peptides derived from exogenous antigens and present them for recognition by CD8-positive cytotoxic T cells. The class II loci are composed of pairs of genes, an A gene and a B gene, which encode an α and β peptide chain, respectively. These dimerize to form the HLA class II molecules, which are expressed only on the surface of specialized antigen-presenting cells, including monocytes, macrophages, and dendritic cells (Figure 30.2(b)).

The DR, DQ, and DP molecules bind to peptides derived from endogenous antigens and present them for recognition by CD4-positive helper T cells (Figure 30.3). The process of antigen presentation is the first step in the activation of a T cell-mediated immune response to the antigen and the HLA molecules therefore play a crucial role in both protection from pathogens and the development of autoimmunity.

The HLA genes are highly polymorphic, some having more than 200 known alleles. This sequence diversity is driven by a strong selective pressure that ensures the recognition of a wide range of antigens to optimize the immune response to a large variety of current and emerging pathogens. As such, the majority of the sequence polymorphism occurs in the gene regions encoding the peptide binding groove. The resulting polymorphic amino acid residues influence the shape and chemical

Figure 30.2  A schematic representation of HLA class I and class II molecules expressed on the cell surface. (a) Class I molecules are composed of an α peptide chain (encoded by an HLA class I gene) and β2 microglobulin (β2m). (b) Class II molecules are heterodimers formed from one α and one β chain, each encoded by a class II locus. Peptides bound in the antigen-binding grooves are denoted by black triangles.
properties of the groove and thereby dictate the repertoire of peptides that can be presented by a given HLA molecule.

Early studies of the HLA gene complex identified disease associations with alleles at individual loci. These genes are in strong linkage disequilibrium (LD), however, which means that recombination between different HLA loci is rare. As a result, combinations of alleles at different genes are inherited together more frequently than expected by chance. When inherited on the same chromosome, this allelic combination is known as a haplotype. For example, the DRB1*0301 allele is most frequently co-inherited with DQA1*0501 and DQB1*0201, forming the DRB1*0301-DQA1*0501-DQB1*0201 haplotype. Haplotypes can only be determined directly in family studies, where transmission of allele combinations from parent to child can be recorded. Strong LD relationships between alleles, however, make it possible to assign haplotypes based on the likelihood of allele co-occurrence, using computer algorithms. Unfortunately such relationships also make it difficult to determine which loci make a genuine contribution to disease risk and which are associated with disease secondary to their coinheritance with other disease markers.

The major HLA-encoded susceptibility determinants for T1DM are the class II DR and DQ genes, although the DP genes and the class I genes also influence disease risk. The effect of a given allele/genotype on the risk of developing disease is generally indicated by the odds ratio (OR), which compares the frequency of the disease occurring in individuals positive for the genetic variant with the frequency in individuals lacking the variant. An OR of 1 indicates that the allele/genotype has no influence on disease risk; typically, an OR value greater than 1 suggests that the marker confers susceptibility to the disease, while markers with OR values less than 1 confer protection.

**HLA-DR and -DQ**

The highest risk of T1DM is conferred by heterozygosity for the DRB1*0301-DQA1*0501-DQB1*0201 and DRB1*04-DQA1*0301-DQB1*0302 haplotypes, referred to as the DR3.DQ2/DR4.DQ8 genotype. This allelic combination is carried by 30–40% of individuals with T1DM, but only around 2.5% of the general population [3]. A recent meta-analysis of multiple ethnic groups suggested that this translates into an OR value greater than 16, an unusually large odds ratio for a complex disease [14]. This is consistent with an earlier study which estimated that the risk of developing T1DM was between 1 in 15 and 1 in 25 among those with the DR3.DQ2/DR4.DQ8 genotype, compared with 1 in 300 in the general population [15]. High risk is also conferred by the DR3.DQ2/DR3.DQ2 and DR4.DQ8/DR4.DQ8 homozygous genotypes (OR = 6.32 and OR = 5.68, respectively, from meta-analysis) [14]. Approximately 95% of subjects with T1DM are positive for DR3.DQ2 and/or DR4.DQ8 compared with 40–50% of the general population [3]. It is noteworthy that heterozygosity for these haplotypes confers a much greater risk than would be expected from the combined OR values observed for each homozygous genotype, suggesting that the alleles on each haplotype interact synergistically to promote disease development. This may be due to the formation of a highly
diabetogenic DQ molecule, encoded in trans by DQA1*0501 on the DR3 haplotype and DQB1*0302 on the DR4 haplotype. This potential synergism supports the conclusion that the HLA genotype of an individual (the combination of alleles inherited from both parents) is a better indicator of disease risk than the presence or absence of specific risk alleles.

Although the DR3.DQ2 and DR4.DQ8 haplotypes are the strongest determinants of T1DM risk, there is a complex hierarchy of allelic, haplotypic, and genotypic effects of the DR/DQ loci, which encompass the spectrum from highly susceptible, through intermediate risk, to very protective markers. The haplotypes most strongly associated with T1DM (predisposing and protective) are summarized in Table 30.1. As shown, the risk conferred by DRB1*0405 and *0401 are associated with the highest risk, while DRB1*0403 confers protection against T1DM [16]. Indeed the protective effect of a given allele from the effects of other alleles with intermediate or very protective effect is not absolute, however, as a small number of individuals with T1DM have been shown to be positive for islet autoimmunity, suggesting that the molecule encoded by DQ6 molecule did not develop autoimmunity to GAD65, even though it is coinherited. Studies of HLA-transgenic animals, which governs the way in which it interacts with autoantigenic peptides and the receptors of autoreactive T cells, shows those most strongly associated with T1DM risk. OR (odds ratio) values are quoted for white Europeans [64], except those marked * which are quoted for Japanese [65].

<table>
<thead>
<tr>
<th>DRB1</th>
<th>DQA1</th>
<th>DQB1</th>
<th>OR</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>0405</td>
<td>0301</td>
<td>0302</td>
<td>11.37</td>
<td>2.71–47.68</td>
</tr>
<tr>
<td>0401</td>
<td>0301</td>
<td>0302</td>
<td>8.39</td>
<td>5.97–11.80</td>
</tr>
<tr>
<td>0301</td>
<td>0501</td>
<td>0201</td>
<td>3.64</td>
<td>2.89–4.58</td>
</tr>
<tr>
<td>0402</td>
<td>0301</td>
<td>0302</td>
<td>3.63</td>
<td>1.76–7.49</td>
</tr>
<tr>
<td>0404</td>
<td>0301</td>
<td>0302</td>
<td>1.59</td>
<td>1.01–2.49</td>
</tr>
<tr>
<td>0802*</td>
<td>0302</td>
<td>5.41</td>
<td>3.29–8.89</td>
<td></td>
</tr>
<tr>
<td>0901*</td>
<td>0303</td>
<td>3.05</td>
<td>2.42–3.83</td>
<td></td>
</tr>
<tr>
<td>0405*</td>
<td>0401</td>
<td>2.33</td>
<td>1.84–2.96</td>
<td></td>
</tr>
</tbody>
</table>

Source: Erlich 2008 [64] and Sugihara 2012 [65]. 95% CI, 95% confidence interval.

dissociate the relative contributions of the DR and DQ genes to disease susceptibility using purely genetic means as the tight LD between the loci hampers attempts to dissociate the effect of a given allele from the effects of other alleles with which it is coinherited. Studies of HLA-transgenic animals, however, have suggested that both DR and DQ molecules play a role in the pathogenesis of T1DM. B10 mice expressing human DR3 (DRB1*03) or DQ8 (DQA1*0301-DQB1*0302), singly or in combination, developed spontaneous T-cell reactivity to glutamic acid decarboxylase (GAD65), a β-cell protein implicated in the autoimmune disease process, while insulitis occurred in DR3/DQ8 double transgenic animals [26,27]. Expression of DQ8 or DR4 (DRB1*0401) in C57BL/6 mice was unable to induce insulitis, but when these animals were crossed with transgenic mice expressing the T cell co-stimulatory molecule, B7.1, on their pancreatic β cells, the resulting double transgenic offspring developed spontaneous diabetes [28,29]. Similarly, the coexpression of B7.1 was necessary for diabetes to develop in mice transgenic for DR3 and/or DQ8 [30]. Interestingly, in this study, the DR3 and DQ8 molecules appeared to be equally permissive for islet autoimmunity. In contrast, Kudva and coworkers reported greater insulitis in the presence of DQ8 alone, with the coexpression of DR3 apparently decreasing the frequency and severity of autoreactivity in double transgenic animals [31]. A similar observation was made in DR4/DQ8 transgenic mice; DQ8 appeared to be the more diabetogenic molecule, with its effect being downregulated by the coexpression of DR4 [29]. These studies suggest a primary role for DQ in the development of autoimmunity, with DR playing a regulatory role. Significantly, transgenic mice expressing the DQ6 molecule did not develop autoimmunity to GAD65, even in the presence of DR3 or DQ8 [26]; this is consistent with the dominant protective effect reported for DQ6.

The influence of a particular HLA molecule on diabetes risk is largely determined by its three-dimensional structure, which governs the way in which it interacts with autoantigenic peptides and the receptors of autoreactive T cells.
The antigen-binding sites of diabetogenic molecules, such as DQ2, DQ8, and I-A^\text{DR7} (the diabetes-predisposing MHC molecule found in nonobese diabetic [NOD] mice), share similar chemical and geometric properties that dictate the shape of the critical pockets into which the amino acid side chains of antigen peptides are anchored during presentation. Likewise there is marked structural similarity between the diabetes-protective molecules, such as DQ6, DR4 (DRB1*0403), and murine I-A^\text{B} [32]. The structural features differ markedly between the two groups, however, leading to differences in antigen selectivity, binding affinity, and molecular stability. It is unclear how this translates into effects on disease risk, but it has been hypothesized that diabetogenic molecules may bind poorly to autoantigenic epitopes, leading to ineffective tolerance induction and subsequent autoimmunity. In contrast, the protective molecules bind well to self antigens, promoting efficient thymic deletion of autoreactive T cells. The potential “loss of function” mechanism for the predisposing molecules is supported by the observation that transgenic overexpression of the diabetogenic allele, I-A^\text{DR7}, in NOD mice protects from diabetes, presumably by compensating for the low affinity of the encoded molecule for autoantigens [33].

Other HLA-encoded susceptibility determinants

The DR and DQ genes cannot completely explain the association between T1DM and the HLA gene region. Teasing out the contribution of other HLA loci, however, is complicated by the high level of linkage disequilibrium within the region, combined with the strong effects of DR and DQ. A number of different approaches have been employed to address this issue, including the analysis of case/control data matched for specific DR-DQ combinations, investigation of the transmission to affected and unaffected offspring of heterozygous markers from parents homozygous for DR-DQ alleles, conditional logistic regression and conditional haplotype analysis. These, and other methods, have identified additional disease risk markers at the HLA-A, -B, -C, and -DPBI loci, in the “class III” gene region and in loci located telomeric and centromeric of the classical HLA gene region (see Figure 30.1).

A number of association studies have implicated variants of the DPBI gene in T1DM risk, but results are conflicting between datasets. A recent analysis of high-resolution genotyping data in a large family-based cohort from the Type 1 Diabetes Genetics Consortium (T1DGC) showed that disease risk was increased by DPBI*0202 and DPBI*0301, while DPBI*0402 conferred protection from diabetes [34]. These associations remained significant after relative predispositional analysis, adjusting for the effects of LD with the DR-DQ loci. The DPBI*0301 allele was found to increase disease risk both in the presence and absence of high-risk susceptibility DR-DQ alleles. In contrast the predisposing effect of DPBI*0202 was seen only in individuals carrying the high-risk and protective DR-DQ haplotypes, while DPBI*0402 conferred protection only in the absence of the high-risk and protective DR-DQ haplotypes.

Similar conditional analyses in the T1DGC dataset have identified HLA class I alleles that influence the risk of diabetes independently of DR and DQ [35,36]. The most significant disease associations reported by Noble and coworkers [35] were with HLA-B*5701 (protective) and B*3906 (predisposing). Increased disease risk was also conferred by A*0201, A*2402, B*1801, and C*0501, while A*1101, A*3201, A*6601, B*0702, B*3502, B*4403, C*0401, and C*1601 were all associated with protection. The predisposing effects of HLA-A*24, B*39, and B*18 and reduced disease risk conferred by A*11 were confirmed by Howson et al. [36]. In addition, HLA-B*13 and B*50 were found to increase T1DM risk independently of DR and DQ, while A*01 conferred protection. It has been postulated that HLA class II molecules play an important role in the initiation of the autoimmune response in T1DM, while class I molecules mediate the later stages of β-cell destruction.

A number of other loci in the MHC have been reported to influence the risk of T1DM independently of DR and DQ, including the TNFA, AIF1, PRSS16, and TTPR3 genes (Figure 30.1), although the association with the latter marker could not be confirmed in an independent study and may be secondary to the effect of DQB1 [14]. Single nucleotide polymorphisms (SNPs) mapping close to LTA and CFB have also been associated with disease on DR3 and DR4 haplotypes.

Overall HLA genes contribute to 40–50% of the genetic risk of T1DM, with the strongest effect seen in individuals developing the disease at an early age. As the incidence of T1DM has risen over past decades, however, a temporal change has been observed in the impact of specific genotypes on disease risk [37]. The proportion of patients with high-risk HLA genotypes has decreased, suggesting that greater environmental pressure is boosting disease penetrance in subjects with moderate- and low-risk HLA genotypes.

Other candidate genes for T1DM

The candidate gene approach has been used to identify a number of T1DM risk loci outside the HLA gene region. This method seeks disease associations with sequence variants at loci thought to play a biologic role in disease etiology. The insulin gene and immune response genes are the most logical candidates for T1DM and a large number of polymorphisms have been reported to be associated with disease risk over the last 30 years. Many of these have proved to be false positive results, artefacts caused by limited sample sizes and population stratification. Some, however, have been replicated in independent datasets and have functional data to support their role in disease etiology. These validated loci are described in the following sections and the role of the immune response genes in T-cell activation is illustrated in Figure 30.3.

The insulin gene (INS)

Outside the HLA locus, the strongest signal for T1DM risk maps within and upstream of the INS gene on chromosome
11p15.5 (OR ~ 2.0). It is unclear whether the etiologic variant is a variable number of tandem repeats (VNTR), located approximately 0.5 kb upstream of INS, or two noncoding SNPs that are in complete LD with the VNTR [38]. Homozygosity for class I VNTR alleles (26–63 repeats) is associated with an increased risk of T1DM, while class III alleles (140–210 repeats) are generally dominantly protective. These markers appear to exert their influence by modulating insulin gene transcription in the thymus. INS expression from class I alleles is more than twofold lower compared with that driven by class III alleles [39]. The resulting low level of insulin in the developing thymus of class I homozygotes is thought to lead to autoimmunity by impairing the deletion of insulin-specific T cells. In contrast, the higher level of insulin present in class III allele carriers promotes T-cell tolerance to the autoantigen and protects against T1DM. In support of this hypothesis, rare class III alleles that completely repress thymic insulin expression confer an increased risk of diabetes [40].

The CTLA4 gene
The CTLA4 gene on chromosome 2q33.2 encodes a transmembrane co-receptor expressed on the surface of T cells. This functions as a negative regulator of T-cell activation via interaction with the B7 molecule on antigen-presenting cells (Figure 30.3). The G allele of the +49A/G SNP in exon 1 of CTLA4 has been implicated as a susceptibility marker for T1DM, but was rejected as a causal SNP by a fine-mapping study, which showed that its effect could be explained by more strongly associated variants in a 6.1 kb noncoding region, 3′ of the gene [41]. These polymorphisms were also associated with susceptibility to Graves’ disease and autoimmune hypothyroidism, suggesting that CTLA4 is a general autoimmunity risk locus. The associated variants were suggested to reduce the production of a secreted, soluble form of CTLA4, which is able to inhibit T-cell proliferation by binding to B7 [41]. This was not replicated in a subsequent study, however, and there is doubt as to whether the 3′ SNPs are directly causal for T1DM [42,43].

The PTPN22 gene
The PTPN22 gene, located on chromosome 1p13, encodes the lymphoid-specific tyrosine phosphatase, LYP. This enzyme dephosphorylates key molecules in the T-cell receptor signaling pathway and hence functions as a negative regulator of T-cell activation (Figure 30.3). Susceptibility to T1DM (and a range of other autoimmune diseases) is associated with the T allele at the nonsynonymous C1858T SNP of PTPN22 (OR = 1.96) [44]. This polymorphism results in an arginine-tryptophan substitution at position 620, an amino acid residue that plays a critical role in the interaction of LYP with its inhibitor, SRC kinase. The gain-of-function polymorphism results in a phosphatase with increased catalytic activity, which is a more potent negative regulator of T-cell activation [45]. This is thought to interfere with the induction of immune tolerance in the thymus or periphery, hence predisposing to autoimmunity.

The IL2RA gene
The interleukin-2 (IL-2) receptor α chain (CD25), encoded by the IL2RA gene on chromosome 10p15, is upregulated in activated effector T cells but constitutively expressed by FOXP3-positive regulatory T cells, which are known to play an important role in self-tolerance. The α subunit greatly enhances the affinity of the IL-2 receptor for its ligand and signaling through IL-2R is crucial for the function of both effector and regulatory T cells (Figure 30.3). Risk of T1DM is associated with three independent haplotypes, spanning a region from 18 kb immediately 5′ of IL2RA to 25 kb into intron 1 of the gene [46–48]. Disease-associated alleles/haplotypes have been shown to influence levels of soluble IL2RA in peripheral blood [47], while a protective haplotype, defined by the rs12722495G allele, correlated with higher expression of CD25 on CD4-positive memory T cells and increased IL-2 responsiveness by these cells [48]. The susceptible haplotype, defined by the A allele of this SNP, was associated with reduced IL-2 responsiveness, decreased expression of FOXP3 by regulatory T cells and a reduction in the ability of Tregs to suppress the proliferation of autologous effector cells [49]. Like CTLA4 and PTPN22, the IL2RA gene appears to be a general autoimmunity locus, with reports of associations in rheumatoid arthritis, multiple sclerosis, and Crohn’s disease.

Genome-wide analyses
Over the last 20 years, advances in high-throughput genotyping techniques and improved understanding of genetic variation in human DNA have enabled researchers to screen the entire genome to identify disease susceptibility loci. Two main approaches have been used: linkage studies and genome-wide association studies.

Linkage studies
Linkage studies identify regions of the genome that are shared more frequently than would be expected by chance by relatives affected by a particular disease. Most studies analyze affected sibling pairs and utilize genetic markers that are scattered throughout the genome at moderate density, typically microsatellites. A significant excess of allele sharing identical-by-descent (IBD) in affected sibpairs suggests that the region containing the marker locus also contains a disease susceptibility locus. The first linkage scan for T1DM identified 20 chromosomal regions with suggestive evidence of linkage to disease, including the HLA and INS gene regions [2]. Subsequent studies have replicated the linkage with HLA, but findings at other loci have been inconsistent. A recent analysis of 2496 multiplex families from the Type 1 Diabetes Genetics Consortium reported significant evidence of linkage to HLA, along with a second locus on chromosome 6q. Suggestive evidence of linkage was observed near the CTLA4 and INS genes, along with two regions on chromosome 19 [50].
Genome-wide association studies (GWAS)

One of the limitations of linkage studies is their inadequate statistical power to detect risk variants with small effect sizes. As our understanding of the genetic basis of complex diseases has grown, it has become apparent that many susceptibility genes have a very modest influence on disease risk (OR ≤ 1.3). Such small effects can be detected more readily by GWAS, provided that the alleles at the risk loci are relatively common in the studied population (minor allele frequency >5%) and sufficiently large datasets are available for analysis. GWAS utilize high-throughput genotyping platforms to analyze several hundred thousand single nucleotide polymorphisms (SNPs), providing much denser coverage of the genome than linkage studies. The analysis relies on the assumption that the SNPs selected for genotyping will “tag” potentially causal variants as a result of linkage disequilibrium. The marker SNPs are genotyped in a discovery cohort of affected (cases) and unaffected, unrelated individuals (controls) and those showing suggestive evidence of association with disease are then taken forward for replication in an independent case/control dataset. Due to the large number of SNPs investigated, very large sample cohorts are required in the replication phase to achieve genome-wide statistical significance for a given variant (widely accepted as a p value less than $5 \times 10^{-8}$). GWAS are therefore generally performed by international collaborative efforts and large consortia, such as the Wellcome Trust Case–control Consortium (WTCCC) and the T1DGC.

Preliminary analysis of 6500 nonsynonymous SNPs in 2029 T1DM cases and 1755 control samples identified a significant disease association with the minor allele of rs1990760 in the IFIH1 gene [51]. This locus encodes interferon-induced helicase C domain 1, which serves as a cellular receptor for double-stranded viral RNA and triggers the production of interferon in response to viral infection. As enteroviruses have been implicated as environmental triggers of T1DM development, the IFIH1 gene may be involved in mediating this effect by activating the innate immune system. The presence or absence of inflammatory signals produced by this system will determine whether T-cell activation in the periphery results in an aggressive effector response or a protective regulatory response.

The first complete GWAS for T1DM was reported by the WTCCC in 2007 [52]. This confirmed associations at known loci (HLA, INS, CTLA4, PTPN22, IL2RA, and IFIH1) and also identified several novel associations with variants on chromosomes 12q24, 12q13, 16p13, 12p13, 5q31, and 4q27. Follow-up studies in different cohorts and meta-analyses of multiple datasets have subsequently confirmed associations at all these loci except 5q31, as well as identifying a large number of additional association signals [11,53]. To date, more than 50 loci have been implicated as determinants of T1DM risk using the GWAS approach; a complete list can be found on the T1DBase website (http://t1dbase.org). Many of these loci are also associated with other autoimmune diseases, suggesting common underlying mechanisms in disease development. The majority of the T1DM risk variants have very small effect sizes, with odds ratios between 1.05 and 1.3, significantly lower than those reported for the HLA genes, INS and PTPN22 (Figure 30.4). Given the statistical power of the GWAS and the large number of SNPs genotyped, however, it is highly unlikely...

---

**Figure 30.4** Odds ratios (OR) of more than 40 T1DM susceptibility loci, identified by candidate gene studies (loci marked *) or genome-wide association studies (data from [53] and [66]). The IFIH1 locus was identified by the analysis of 6500 nonsynonymous SNPs.
that any additional common variants with large effects on disease risk were missed.

It is important to note that, like linkage analyses, GWAS do not directly identify causal variants. The designations often applied to associated variants can therefore be misleading. For convenience, signal SNPs are generally assigned the name of the closest gene, or the most plausible candidate gene in close proximity, as a reference point for their location, although some associated variants map to apparent gene deserts, with no known annotated genes. Close physical proximity of a candidate gene, however, does not necessarily mean that the gene has any functional involvement in disease pathogenesis or that it accounts for the risk associated with the nearby variant. Many disease associated variants map to large blocks of LD, which may span hundreds of kilobases and encompass many genes. In most cases the signal SNP will merely be a marker for a causal variant located somewhere in the LD block. Fine mapping is therefore required to narrow down the likely location of this etiologic SNP. This is generally achieved by deep sequencing of the LD region of interest to identify all the polymorphisms that might contribute to the observed association. These are then genotyped in large case–control cohorts and conditional regression analysis is used to tease out the contribution of each variant and determine whether this can be explained by LD with other SNPs or is an independent phenomenon. This approach often narrows down disease association to haplotypes comprising many closely correlated variants, but further refinement cannot be achieved using genetic means due to very strong LD between the SNPs. Functional analysis is then necessary to home in on the causal variant(s). This may be accomplished by identifying polymorphisms known to influence gene expression levels (from publicly available expression QTL (eQTL) databases) or missense variants within the associated haplotype that could disrupt protein function.

The T1DM risk loci identified by GWAS contain over 300 protein-encoding genes, many of which are good candidate genes due to the role of their products in immunologic or metabolic pathways. This estimate excludes genes located outside the associated LD blocks that could have regulatory elements within the blocks. As functional SNPs may exert their effects via long-distance gene regulation, their contribution to disease risk may be mediated by genes located several kilobases away from the LD blocks in which the SNPs lie. This highlights the difficulties involved in conclusively identifying the causal gene variants and elucidating the genetic mechanisms underlying T1DM risk. To date, only the HLA gene region, INS and IL2RA have been fine-mapped with any degree of thoroughness, although functional studies also support the PTPN22 R620W variant as a causal SNP.

To date, GWAS for T1DM have been performed exclusively in white European datasets, due to the difficulty in recruiting sufficiently large cohorts from other ethnic groups, in which the disease is relatively rare. One drawback of this limited approach is that some susceptibility loci might have been overlooked if their minor alleles occur with low frequency (<5%) in white Europeans, as the studies would have been underpowered to detect the associations. Allele frequencies at some loci are increased in certain populations as a result of population genetic drift, thus boosting statistical power to detect disease associations with these variants. It is therefore possible that additional susceptibility loci for T1DM may be discovered if different ethnic groups are analyzed by GWAS. A second issue is that it is not known whether the T1DM risk loci identified in white Europeans have global relevance. To date, studies have not attempted to replicate associations at all the GWAS loci in non-White populations as available datasets are too small. Candidate gene studies have shown suggestive evidence of population-specific effects, however. For example, the M55V variant of the SUMO4 gene is strongly and consistently associated with T1DM in East Asian populations, but not in white Europeans [53,54]. In contrast, the disease association with the R620W variant of PTPN22, reported in white Europeans, is not seen in Asians [55], perhaps due to the rarity of the minor allele in the latter population. Genetic heterogeneity between different populations is perhaps to be expected, given that most of the genes influencing T1DM risk are likely to act via the immune system. Population survival is dependent upon the adaptation of the immune response to local environmental insults, which are likely to differ in different areas of the world. This shaping of the immune repertoire might be reflected in ethnic differences in immune response genes. As a result, it may not be possible to extrapolate all genetic associations observed in European populations to populations of different racial ancestry. There is clearly a need to broaden genome-wide analyses to encompass a more diverse set of populations of different ethnic ancestry to fully understand the contribution of genetic factors to disease risk on a global scale. Exploitation of differences in LD patterns in different ethnic groups may also facilitate the fine-mapping of causal variants for T1DM.

Other genetic markers— the missing heritability”

Collectively the confirmed T1DM risk loci account for approximately 70% of disease heritability, with around 40–50% being attributed to the HLA genes. These figures are well in excess of the 10–20% of heritability of other complex diseases that can be explained by genetic factors. Experience from GWAS suggests that overall disease risk is likely to be influenced by many genes, most having a weak biologic effect. This may be due to the subtle effects of risk alleles on gene function or the modest contribution of individual gene products to the biologic pathways involved in disease pathogenesis. None of the confirmed T1DM risk variants have complete penetrance and are therefore neither necessary nor sufficient for disease to develop. This makes it difficult to use genetic profiling to predict disease risk as T1DM can develop in the absence of susceptibility variants and does not always occur in subjects with known risk markers. Furthermore the combination of susceptibility variants underlying T1DM
Molecular genetics of type 1 diabetes

Clinical utility of T1DM susceptibility genes

One of the goals of genetic studies of complex disease is to identify a profile of susceptibility variants that can be used to predict an individual’s risk of developing a given disease. The long prodrome for T1DM, characterized by progressive loss of \( \beta \)-cell mass, provides an attractive opportunity for intervention to prevent disease development if “at-risk” individuals can be identified. Currently the best markers of disease risk are a positive family history of T1DM and the presence of autoimmune antibodies to islet cell proteins. Over 90% of T1DM patients have no affected relatives, however, so effective preventive efforts will need to target the general population. Unfortunately screening such huge numbers of individuals for islet autoantibodies is logistically unfeasible, particularly given the need for repeated annual testing for those with a negative result. Genetic markers are therefore needed to stratify risk in the general population, to significantly decrease the target population for immunologic screening while still capturing the majority of future cases of T1DM. HLA class II genes have been widely used in research studies to identify high-risk individuals (see later), but testing for HLA alleles alone lacks specificity, sensitivity, and positive predictive value, limiting its use in disease prediction at a population level. Incorporating genotype data for susceptibility SNPs outside the MHC improves prediction slightly but it has been suggested that even if all risk loci were discovered and included in the prediction model, the achievable predictive power might still be inadequate to reliably select individuals from the general population for targeted intervention [58].

Although the HLA genes may have limited utility in large-scale population screening for disease prevention, they are useful markers for stratifying disease risk in research studies. Birth cohorts examining the natural history of T1DM and the impact of environmental determinants on disease development, such as Diabetes Autoimmunity Study in the Young (DAISY), BABYDIAB, and The Environmental Determinants of Diabetes in the Young (TEDDY), use HLA typing to identify newborns in the general population with high and moderate genetic risk of disease and to stratify risk in children with a family history of T1DM [15,59,60]. High genetic risk is defined by the \( DR3.DQ2/DR4.DQ8 \) genotype, while moderate risk is conferred by genotypes carrying one of these high risk haplotypes, in the absence of \( DR15.DQ6 \). Children positive for \( DR3.DQ2/DR4.DQ8 \) have a 5% risk of developing T1DM by the age of 15 years [14]. If they also have an HLA-identical sibling with T1DM, the risk is even higher; a 65% risk of developing islet autoimmunity by the age of 7 years and a 55% risk of developing T1DM by the age of 12 years [61]. In contrast, the presence of the highly protective \( DR15.DQ6 \) haplotype markedly reduces disease risk, even among individuals with a family history of disease and evidence of islet autoimmunity [20]. HLA genes are also valuable tools for selecting high-risk individuals for trials of novel preventive strategies for T1DM, such as the PrePOINT trial of autoantigen vaccination with oral insulin (http://www.diabetes-point.org) and the TRIGR study of dietary intervention in infancy (http://trigr.epi.usf.edu).

One of the major benefits of the genetic studies of T1DM is the improved understanding of the immune dysregulation underlying the disease. Both innate and T cell-mediated adaptive immunity are implicated by genetic association studies and functional analyses of disease susceptibility variants suggest
that impairment of autoantigen-driven TCR signaling and inefficient negative selection of autoreactive T cells are important mechanisms in the development of autoimmunity. To date, the majority of functional studies have been performed on candidate genes but extension of this approach to the large number of loci identified by GWAS will likely yield further useful information on the molecular pathways involved in disease development and provide valuable insight into the mechanisms underlying T-cell tolerance. Such information may identify novel therapeutic targets for the prevention of T1DM. As shown by PPARγ in type 2 diabetes, a gene does not have to have a large effect on disease risk to be a good target for drug treatment. In this vein, there is currently considerable interest in developing inhibitors of LYP, to treat a range of autoimmune diseases [62], and compounds able to manipulate thymic insulin expression, which would have therapeutic potential in the treatment of T1DM [63].

**Conclusion**

Although considerable progress has been made in recent years to further our understanding of the genetic basis of type 1 diabetes, there is still a substantial proportion of inherited susceptibility that is, as yet, unexplained. Furthermore the causal variants at most of the validated risk loci remain to be identified, as do the mechanisms by which they influence disease risk. Future research aiming to address these gaps in our knowledge should integrate multiple approaches, including sequence-based fine mapping, whole-exome/whole-genome sequencing to identify rare variants, transcriptomics and mechanistic studies, gene network and pathway analysis, epigenetic regulation of gene function and the analysis of gene–gene and gene–environment interactions. Such studies may identify new therapeutic targets for T1DM and/or markers of extreme genetic risk that can be utilized in targeted primary prevention.

**References**


52 Wellcome Trust Case Control Consortium: Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007;447:661–678.


SECTION VI

Obesity
CHAPTER 31

The relationship between obesity and type 2 diabetes—the role of gut factors

Daniel S. Hsia and William T. Cefalu
Pennington Biomedical Research Center, Louisiana State University, Baton Rouge, LA, USA

Key points

- There are two key pathophysiologic defects underlying the progression from normal glycemic states to pre-diabetes and to type 2 diabetes: impaired insulin action due to insulin resistance and impaired β-cell function/insulin secretion.
- Gut hormones have important roles in the regulation of appetite, food intake, and insulin secretion.
- Glucagon-like peptide-1 (GLP-1) is released into the circulation after a meal and has various physiologic effects, including increasing insulin secretion, suppressing glucagon secretion, delaying gastric emptying, and suppressing appetite.
- Ghrelin is a 28-amino acid peptide hormone produced predominantly in the stomach and is the only known orexigenic gut hormone identified to date.
- The human gut microbiota contributes to multiple host processes including the defense against pathogens at the gut level, immunity (mediated through a number of signal molecules and metabolites), energy harvest, the development of the intestinal microvilli, maintenance of the intestinal barrier, and the synthesis of several vitamins.
- The gut microbiota’s role in obesity and diabetes is still evolving, but likely involves differences in the diversity and composition of the gut microbiota among individuals leading to pathophysiologic changes such as varied energy harvest from the diet and low-grade chronic inflammation (due to intestinal permeability) which may contribute to insulin resistance.
- Unlike the human genome, the composition of the intestinal microbiota can be modified and may depend on specific dietary habits.

Introduction

The rise in obesity has been mirrored by the increase in the incidence of type 2 diabetes (T2DM). It is estimated that 60–90% of all people with T2DM are obese [1]. However, the majority of obese people do not develop diabetes, and thus obesity is only one factor that leads to the development of T2DM. There are several factors that have been recognized as major components of the obesity–diabetes relationship including insulin resistance, pancreatic dysfunction, and increased hepatic glucose production. Insulin resistance and relative insulin deficiency have been the factors most studied. In the early stages of insulin resistance, insulin secretion increases to compensate for defects in insulin action. Diabetes occurs when the insulin secretory capacity can no longer compensate for decreased glucose uptake due to insulin resistance [1]. There are both genetic and environmental contributions to the development of obesity and diabetes, and their rising prevalence necessitates further studies on modifiable factors and novel treatment options. A family history of diabetes puts an individual at higher risk for developing diabetes. However, studies such as the Diabetes Prevention Program demonstrated that lifestyle interventions including weight loss and exercise could reduce the incidence of diabetes in people at high risk [2]. In addition, with the introduction of incretin-based therapy for T2DM, it is clear that modulation of gut hormones can also affect glycemia. Gut hormones have important roles in the regulation of appetite, food intake, and insulin secretion. There is also an emerging role of gut microbiota on human health. The gut microbiota’s role likely involves differences in the diversity and composition of the gut microbiota among individuals leading to pathophysiologic changes such as varied energy harvest from the diet and low-grade chronic inflammation (due to intestinal permeability) which may contribute to insulin resistance.

In this chapter, the focus will be on the role of gut factors, specifically gut hormones and the gut microbiota, which contribute to the pathophysiology of obesity and T2DM.
The relationship of obesity to insulin resistance and diabetes

The World Health Organization has deemed the increased prevalence of obesity and diabetes as a “twenty-first century epidemic” [3]. More than half of the world’s population is considered overweight and being overweight is associated with several comorbidities such as T2DM, cardiovascular diseases, hypertension, dyslipidemia, respiratory diseases, osteoarthritis, and depression [3]. According to Ford et al., for every kilogram of weight gain, the risk of diabetes increases between 4.5 and 9% [4]. The relative risk for an obese individual to develop T2DM is 10-fold for women and 11.2-fold for men [5]. The most critical factor in the emergence of metabolic diseases is obesity.

One of the causal links between obesity and T2DM is the development of insulin resistance. Adipose tissue modulates metabolism by releasing free fatty acids (FFAs) and glycerol, hormones (including leptin and adiponectin), and pro-inflammatory cytokines [6]. It is well known that insulin sensitivity decreases with advancing age and is correlated with increasing body fat content. FFAs are an important link between obesity and insulin resistance. Acute elevations of plasma FFAs have been shown to increase insulin resistance in a dose-dependent manner in both diabetic and nondiabetic individuals [3]. Conversely, insulin-mediated glucose uptake and glucose tolerance improve with an acute decrease in FFAs after treatment with the antilipolytic agent acipimox [7]. Moreover, visceral and abdominal fat have been shown to have a negative correlation with insulin sensitivity [8] and thus have a prominent role in the pathogenesis of insulin resistance. Visceral fat cells are less sensitive to suppression of lipolysis by insulin [9]. FFAs produced by visceral adipocytes and entering via the portal circulation in the liver can induce hepatic insulin resistance, particularly by enhancing gluconeogenesis [3]. Gastaldelli et al. demonstrated that in patients with T2DM, visceral fat accumulation (independent of BMI) has a negative impact on glycemia through a decrease in peripheral insulin sensitivity and an enhancement of gluconeogenesis [10]. In muscles, the chronic decrease in glucose utilization due to the preferential use of lipids as an energy source limits the use of glucose from glycogen stores and inhibits glycogen synthase activity.

Ultimately, factors associated with obesity may lead to development of T2DM. There are two defects underlying this progression: impaired insulin action due to insulin resistance and impaired β-cell function/insulin secretion. In the early stages of insulin resistance, β cells in the pancreas secrete increased amounts of insulin to maintain euglycemia, thereby compensating for attenuated insulin action. However, after years of obesity and due to functional defects in insulin secretion, the β cells cannot maintain such high rates of insulin secretion, and the “relative” insulin deficiency given the degree of insulin resistance leads to impaired glucose tolerance and eventually T2DM [3,9]. This progression is illustrated in Figure 31.1 with obese diabetic individuals having the highest glucose and the lowest insulin levels.

---

Figure 31.1 Plasma glucose and insulin concentration in the course of a 3-h oral glucose tolerance test in four groups of obese patients and lean control subjects, young (left) and older (right). Source: Felber JP, Golay A. Int J Obes Relat Metab Disord. 2002 Sep; 26 Suppl 2:S39 – 45. Reproduced with permission of Nature Publishing Group.
Gut hormone regulation of appetite, food intake, and insulin secretion

Before we can understand the role of gut hormones in the pathophysiology and treatment of diabetes, we need to understand their major role in appetite, food intake, and insulin secretion. The gastrointestinal (GI) tract is the largest endocrine organ in the body and is believed to have an important role as a source of various appetite-regulating peptide hormones [11]. These hormones are thought to have a critical role in initiating and terminating eating behaviors. Postprandial satiety is believed to be regulated by communication between the gut and appetite-regulating centers in the brain, with the hypothalamus responsible for nutrient and energy sensing and corresponding adjustments in food intake [12]. This section will focus on the broad categories of pancreatic polypeptide-fold peptides and incretins.

Pancreatic polypeptide (PP)-fold peptides

The pancreatic polypeptide (PP)-fold family consists of neuropeptide Y (NPY), and the two gut hormones peptide tyroisine-tyroisine (PYY) and pancreatic polypeptide (PP); each shares a hair-pinfold motif necessary for receptor binding and has a distinct role in appetite regulation [13].

Peptide YY (PYY)

PYY is a full-length 36-amino acid peptide that is made and released from the L cells of the GI tract. PYY is processed in the circulation by dipeptidyl peptidase 4 (DPP-4), releasing the first two amino-acid residues and creating the active and dominant peptide PYY3-36[14]. PYY3-36 appears to act in the hypothalamus as a satiety signal and has been shown to have anorexigenic effects in both normal weight and obese individuals. Circulating levels of PYY3-36 are influenced by meal composition and calorie content, and become elevated within 1 hour after feeding [12]. PYY3-36 infusion was found to reduce calorie intake similarly in obese subjects and in lean controls (30% vs. 31%) [15]. Since the anorexigenic effect of exogenous PYY3-36 is fully intact in obese individuals, PYY resistance is not thought to exist. This has encouraged longer-term weight loss studies involving chronic administration of PYY3-36 [12]. Significant increases in circulating PYY3-36 levels have also been reported post-GI surgery, possibly contributing to the initial and long-term maintenance of weight loss attributed to the procedure [14].

Pancreatic polypeptide (PP)

PP is a 36-amino acid anorexigenic peptide primarily made and released from the endocrine pancreas, and to a lesser extent, from the colon and rectum. PP is regarded as part of the “ileal brake,” which slows the transit of food through the gut. PP slows gastric emptying, inhibits pancreatic exocrine secretion and reduces gallbladder contraction [16]. The main stimulus of PP release is the ingestion of food. In humans, postprandial levels of PP are proportional to calorie intake and levels remain elevated for approximately 6 h after eating [13]. PP release is inversely correlated to body weight, with obese subjects displaying reduced postprandial PP secretion, and anorexic individuals displaying elevated postprandial levels [13]. Peripheral PP administration leads to an increase in energy expenditure and a reduction in body weight in rodents in addition to a reduction in appetite and food intake in both lean and obese humans [12]. Although the exact mechanism of PP’s effect is still unknown, these data support its potential as an anti-obesity therapy.

Incretins

Incretins are gut-derived factors that stimulate insulin release and are produced by selective posttranslational cleavage of proglucagon, a 160-residue peptide, expressed in the α-cells of the pancreas and L cells of the intestine and the CNS [13].

Glucagon-like peptide (GLP-1)

GLP-1 is a 29-amino acid polypeptide secreted from the L cells of the distal gut and is a cleavage product of proglucagon. The two synthesized forms of GLP-1 are GLP-17-37 and GLP13-36amide, with further cleavage required to produce the two bioactive fragments, GLP-17-26amide and GLP-17-37 [17]. GLP-1 is released into the circulation after a meal and has various physiologic effects, including increasing insulin secretion, suppressing glucagon secretion, delaying gastric emptying, and suppressing appetite [14]. In addition, in preclinical studies, GLP-1 has been found to stimulate β-cell proliferation, to enhance the differentiation of new β cells in the pancreatic duct epithelium, and to inhibit apoptosis of β cells.

GLP-1 and GIP are responsible for what is known as “the incretin effect,” which is the enhanced secretion of insulin in response to oral administration of glucose compared with intravenous administration of glucose [18]. GLP-1 is released in the gut in proportion to ingested calories. In both lean and obese humans, peripherally administered GLP-1 has been shown to exert anorexigenic effects, with other possible influences on food intake being linked to a reduction in gastric emptying and a suppression of gastric acid secretion. Obese individuals have been reported to elicit delays in the postprandial release of GLP-1, and thus present with reduced circulating levels of the peptide. Because of inactivation and clearance by the enzyme DPP-4, the half-life of GLP-1 is an estimated 5 min, thus presenting a major hurdle to its possible therapeutic utility. Currently, approaches seek to take advantage of GLP-1 effects by either extending the half-life of GLP-1 with endogenous inhibition of the responsible enzyme or providing a molecule that resists the degradation. Such pharmacologic approaches are realized with the development of GLP-1 analogues and DPP-4 inhibitors. Inhibition of DPP-4 has had useful applications in the treatment of T2DM, but less promising results have been demonstrated in terms of its anti-obesity utility.
Oxyntomodulin (OXM)

OXM shares the same precursor molecule as GLP-1, is co-secreted with GLP-1 from the L cells of the distal small bowel following feeding, and its release is proportional to the meal calorie content [14]. OXM inhibits gastric acid secretion and gastric emptying. OXM acts as a dual agonist at the GLP-1 and glucagon receptors [14]. GLP-1 and glucagon are known to suppress food intake and glucagon can increase energy expenditure. This combination of decreased food intake and increased energy expenditure would be expected to enhance weight loss but no OXM analogues have yet to be approved for human use. Similar to GLP-1, the potential therapeutic utility of OXM may, in part, be hindered due to its inactivation by DPP-4. Although OXM has a less potent incretin effect compared to GLP-1, the effects of OXM on food intake in humans are more potent with reportedly less nausea than GLP-1 [12].

Glucose-dependent insulinotropic polypeptide (GIP)

Initially called gastric inhibitory polypeptide due to its inhibitory effects on gastric acid secretion in dogs, this 42-amino acid polypeptide was renamed glucose-dependent insulinotropic polypeptide to reflect its role in glucose metabolism [19]. GIP is secreted from K cells after eating a meal and is inactivated rapidly by DPP-4. Along with GLP-1, GIP acts as an incretin to potentiate glucose-stimulated insulin release. GIP has direct anabolic effects on adipose tissue, including stimulation of glucose import, fatty acid synthesis, lipogenesis, and inhibition of lipolysis [14]. GIP indirectly influences energy balance and body weight due to its effects on lipid metabolism; however, direct infusion of GIP into healthy male subjects does not affect appetite, energy intake, or energy expenditure [20].

Cholecystokinin (CCK)

CCK was the first gut hormone shown to affect appetite. Levels of CCK in the plasma rise within 15 min of meal initiation, and CCK reduces food intake in a dose-dependent manner [12,21]. CCK is predominantly produced and released by the L cells of the duodenum and jejunum. It exists in several bioactive forms including CCK-8, CCK-22, CCK-33, and CCK-58; the numerical suffix denotes the number of amino acids [21]. CCK has several roles including: stimulation of gallbladder contraction, pancreatic enzyme secretion, and inhibition of gastric emptying. CCK is also present in the hypothalamus, predominantly in the median eminence and ventromedial nucleus, and is the most abundant neuropeptide in the CNS [12]. Administration of CCK has been shown to reduce food intake in both rodents and humans, which has led to its investigation as a therapeutic target for obesity. However, the increase in meal frequency, tolerance after infusion, and its short half-life have limited CCK’s therapeutic utility [12].

Ghrelin

Ghrelin is a 28-amino acid peptide hormone produced predominantly in the stomach and is the only known orexigenic gut hormone identified to date. Ghrelin binds to the growth hormone secretagogue receptor and is also expressed in the NPY neurons of the hypothalamus. Circulating ghrelin levels increase with fasting, encouraging food intake, and then decrease after a meal [14]. Administration of ghrelin has been shown to induce appetite and food intake, suggesting it may be useful therapeutically to reverse pathologic states of cachexia [13]. Plasma ghrelin levels correlate inversely with body weight, and diet-induced weight loss increases circulating levels of ghrelin in obese individuals [22]. Also, the expected postprandial fall in circulating ghrelin levels is attenuated, or even absent in obese people, suggesting that ghrelin has a role in the pathophysiology of obesity [12]. Pharmacologic blockade of ghrelin results in decreased food intake and body weight in rodents, and ghrelin- or ghrelin receptor-deficient rodents are resistant to diet-induced obesity [12]. Thus, ghrelin antagonism may be a promising strategy to treat obesity.

Other related hormones

Amylin

Islet amyloid polypeptide (IAPP), also known as amylin, is a 37-amino acid peptide synthesized in pancreatic β cells that is co-secreted with insulin in response to nutrient intake. It is also produced in smaller amounts by enteroendocrine cells and enteric neurons within the gastrointestinal tract [23]. Amylin is briefly mentioned here because of its role in glucose regulation and food intake. Amylin decreases postprandial blood glucose levels through inhibition of gastric emptying and suppression of glucagon secretion. Activation of amylin receptors in the area postrema in the brain induces satiety and leads to decreased food intake [23].

Role of gut hormones in obesity and type 2 diabetes

There is increasing evidence that the function and efficacy of gut hormones change in the obese individual [19]. Fasting CCK levels as well as the normal postprandial rise in CCK are lower in morbidly obese subjects compared with either obese or lean controls [24]. As for PYY3-36, fasting levels are inversely proportional to body mass and the postprandial rise in PYY3-36 is blunted in obesity. Obese humans have been reported as needing to consume twice as many calories to achieve the same plasma PYY3-36 levels as lean controls [25]. Therefore, PYY3-36 deficiency has been proposed to be involved in the maintenance of obesity by reducing satiety. While basal ghrelin levels are lower in obese compared to lean individuals [26], food intake is higher in obese compared to normal weight individuals when low doses of ghrelin are given, indicating that obesity increases ghrelin sensitivity [27]. Incretin levels also change in obesity. Fasting GLP-1 is reduced, and its response to an oral glucose load is blunted [19]. The reduction in GLP-1 may contribute to the impaired glucose tolerance that occurs in obesity. Changes in GIP levels have not been consistently observed in obese individuals; however, in theory increased
GIP levels may propagate obesity by increasing fatty acid uptake into adipose tissue [19].

Studies by Nauck et al. have shown that the incretin effect is severely reduced or lost even in relatively lean type 2 diabetes patients (BMI ~25 kg·m⁻²) [28]. Knop et al. demonstrated a similar impairment in the incretin effect in individuals with T2DM as well as those with diabetes secondary to chronic pancreatitis, indicating that an impaired incretin effect is likely a consequence of developing diabetes rather than a cause of T2DM [29]. Interestingly, infusion of large amounts of GLP-1 resulted in near normal insulin responses in individuals with T2DM whereas infusion of GIP has no significant effect [30]. This has led to the development of DPP-4-resistant analogues (such as exenatide and liraglutide) as well as inhibitors of DPP-4 (such as saxagliptin, sitagliptin, and vildagliptin) to protect GLP-1 from degradation and to augment its insulinotropic activity [31–33]. New approaches are seeking to extend the half-life of GLP-1 with the development of more stable GLP-1 analogues such as extended release exenatide and albiglutide.

The role of microbiota in obesity and diabetes

Since the announcement of the Human Microbiome Project [34] in 2007, relationships between gut bacteria and a variety of health conditions have been reported. Staying with the gut theme for this chapter, we switch our focus to another emerging contributor to the pathophysiology of obesity and diabetes: gut microbiota.

Gut microbiota in health and disease

The human fetus is microbiologically sterile and is colonized by bacteria from the mother and the surrounding environment at birth. The initial microbiota is relatively unstable and undergoes dramatic changes before stabilizing around weaning [35]. The human gastrointestinal (GI) tract contains a complex consortia of trillions of microorganisms (approximately 1 x 10¹³ to 1 x 10¹⁴, biomass > 1 kg), thousands of bacterial phylotypes, as well as hydrogen-consuming methane-producing microorganisms with a collective genome (termed the microbiome), the majority of which resides in the colon [36]. As a whole, the microbiome represents more than 100 times the nucleotide content of the human genome. Thus, the human “metagenome” consists of a mixture of genes embedded in the human genome and in the genomes of our microbial partners. The microbiota and its microbiome provide humans with additional gene products that we lack, and these serve many functions in maintaining our bodies’ homeostasis. The human gut microbiota may be regarded as a “microbial organ” within the gut, which contributes to multiple host processes including the defense against pathogens at the gut level, immunity (mediated through a number of signal molecules and metabolites), energy harvest, the development of the intestinal microvilli, maintenance of the intestinal barrier, and the synthesis of several vitamins. Moreover, accumulating evidence indicates that the gut microbiota plays a crucial role in conditions such as obesity, diabetes, nonalcoholic fatty liver disease, inflammatory bowel disease (IBD), and even cancer [36].

Gut microbial composition among healthy humans is influenced by host genotype, diet, age, and gender; and it appears that a variety of diseases and drugs can modulate microbiome composition and activities.

In spite of the large microbial diversity in humans, there are only a small number of microbial phyla that are numerically dominant: Firmicutes (~60%), Bacteroidetes (~15%), Actinobacteria (~15%), Verrucomicrobia (~2%), Proteobacteria (~1%), and Methanobacteriales (~1%) [36]. Interestingly, adult monozygotic and dizygotic twins may have a similar microbiota even if they live in different locations [37]. These findings suggest that a shared environment early in life and the maternal inoculum have a large impact upon the gut microbiota in adulthood.

The relationship of gut microbiota to energy balance, obesity, and obesity-related disorders

Gut microbiota exert a crucial role in the development of fat mass and altered energy homeostasis. The earliest evidence supporting this hypothesis came from a study showing that germ-free mice (mice raised in the absence of any microorganisms) are leaner compared to mice with microbiota since birth. Importantly, the colonization of germ-free mice with a gut microbiota caused an increase in fat mass and insulin resistance. This study showed that the gut microbiota constitute a novel environmental factor that regulates fat storage [38]. Moreover, Ridaura et al. demonstrated that an increased adiposity phenotype was transmissible in mice via an ingested fecal sample from an obese human. Co-housing these obese mice with lean mice prevented the increased adiposity phenotype when the mice were fed a low saturated fat, high fruits and vegetables diet. Rescue correlated with invasion of members of the Bacteroidetes species from the microbiota of the lean mice into the microbiota of the obese mice. These findings reveal transmissible, rapid, and modifiable effects of diet-by-microbiota interactions [39].

Obesity is recognized to be associated with changes in the gut microbiota diversity and composition. In addition, individuals with a low bacterial richness/low gene count (<480,000 genes) in their microbiota have more overall adiposity, insulin resistance, dyslipidemia, and a more pronounced inflammatory phenotype when compared with individuals with high bacterial richness [40]. Various studies have characterized the gut microbiota in mouse models of obesity, with many of the results highlighting an increase in Firmicutes and a decrease in Bacteroidetes associated with obesity [41,42]. Early gut microbial composition has been shown to be predictive of the development of obesity in a variety of populations: small children [43], obese adolescents [44], and pregnant women [45]. However, most of these studies have been performed in a limited number of subjects, using microbial methods that were not state-of-the-art according to current standards, and were mostly retrospective in design. In addition, a study by Duncan et al.
found no relationship between obesity and the proportions of Bacteroides and Firmicutes in obese men [46]. Schwertz et al. determined lower ratios of Firmicutes to Bacteroidetes in overweight human adults compared to lean controls [47]. Thus in humans, the increase in the Firmicutes/Bacteroidetes ratio as it relates to obesity continues to be one of active study and it is likely that our understanding of what specific microbial groups contribute to obesity risk will continue to evolve.

In terms of T2DM, Larsen et al. demonstrated a significantly reduced amount of Firmicutes in people with type 2 diabetes compared with a control group. Moreover, the ratio of Bacteroidetes to Firmicutes positively correlated with plasma glucose concentration but not with body mass index [48]. Comparisons of gut microbiota between obese and lean individuals with or without T2DM revealed that people with type 2 diabetes had reduced levels of Faecalibacterium prausnitzii, a member of the phylum Firmicutes. This reduction correlated with increased levels of inflammatory markers and suggests that F. prausnitzii potentially functions as a probiotic to alleviate insulin resistance [49].

Potential mechanisms of disease related to gut microbiota

There are multiple features of the gut microbiota that can promote obesity and insulin resistance (Figure 31.2). One of the proposed mechanisms of how the gut microbiota influences obesity is that the gut microbiota have the capacity to increase the energy harvested from the diet. Increased energy harvest correlates with increased fermentation capacity of the microbiota and increased levels of short chain fatty acids (SCFAs) [35]. SCFAs are produced by the colonic microbiota through anaerobic fermentation. The chief substrates for colonic fermentation are undigested carbohydrates, namely dietary fiber and resistant starch [50]. Acetate, propionate, and butyrate are the main SCFAs produced by the microbiota in the colon [51]. Normal colonic epithelia derive 60–70% of their energy from SCFAs, particularly butyrate. Propionate is largely taken up by the liver and is a precursor for gluconeogenesis, liponeogenesis and protein synthesis. Acetate enters the peripheral circulation and is a substrate for cholesterol synthesis. Analysis of fecal SCFA concentration in a group of lean and obese subjects showed significantly higher total SCFA content in obese subjects [47]. Another smaller study of 11 lean and 11 overweight or obese individuals showed a positive correlation between fecal SCFAs and Firmicutes abundance. Fecal SCFAs were higher in the overweight/obese group despite similar dietary intakes [50]. Studies have shown that mice deficient in either of the SCFA receptors, Gpr41 and Gpr43, are leaner than their wild-type counterparts, further implicating SCFAs in the development of obesity [35].

Studies have demonstrated that the gut microbiota can also trigger inflammation, thereby possibly playing a major role in the onset of insulin resistance and T2DM [41]. Gut microbiota-derived lipopolysaccharide (LPS) is a key molecule involved in the early development of inflammation and metabolic diseases. LPS is a powerful pro-inflammatory molecule from the cell wall of Gram-negative bacteria and is continuously produced in the host gut with the death of Gram-negative bacteria [41]. High-fat feeding has been shown to not only increase fatty acid levels but to also increase circulating LPS levels [35]. Direct evidence for the role of LPS was demonstrated by chronic infusion of LPS over a 4-week period in chow-fed mice, which led to increased adiposity, increased macrophage infiltration in adipose tissue, hepatic inflammation, and hepatic insulin resistance [52].

Emerging evidence suggests that gut barrier disruption could be responsible for the metabolic endotoxemia present in obesity and T2DM, as has been shown particularly in models of Crohn’s disease [41]. Conversely, modulating gut microbiota to reduce endotoxemia may improve diabetes risk factors. For instance, an increase in the gut barrier-protective Bifidobacteria improves high-fat-diet-induced diabetes in mice through a mechanism associated with decreased endotoxemia [53]. Studies have demonstrated that the increase in gut permeability observed in obese mice may be associated with an alteration in the expression, localization, and distribution of two tight-junction proteins (occludin and zonula occludens-1) in the small intestine [41]. Moreover, LPS is taken up from the gut lumen either via a chylomicron-dependent mechanism and/or through tight junctions. The gut microbiota regulates gut permeability by at least two mechanisms—via either GLP-2 or the endocannabinoid system [35]. Treatment of obese ob/ob mice with prebiotics induces endogenous GLP-2 levels that are associated with improved gut permeability and reduced plasma LPS levels. Pharmacologic inhibition of GLP-2 signaling negated the effects of prebiotics, establishing a direct link between GLP-2 and gut permeability [54]. Prebiotics can reduce colonic expression of the endocannabinoid receptor CB1 in obese mice, and treatment with a CB1 receptor antagonist prevented degradation of tight-junction proteins and reduced plasma LPS levels [55].

Intestinal alkaline phosphatase (IAP) is known to be involved in the breakdown of dietary lipids and also plays an important role in LPS detoxification by dephosphorylating the lipid portion of LPS. Obesity appears to be associated with a decrease in IAP activity. Thus, modifications of the IAP activity by a high-fat diet, gut microbiota, or other compounds may contribute to the alterations in the gut barrier functions in obesity and T2DM. Moreover, an increase in IAP activity is associated with a reduction in metabolic endotoxemia [41].

In a recent study, Vrieze et al. demonstrated that infusion of intestinal microbiota from lean healthy donors temporarily improves insulin sensitivity in individuals with metabolic syndrome [56]. Although this link is well established in a mouse model of obesity, more studies are needed to confirm the involvement of gut microbiota alterations and gut barrier functions in the metabolic endotoxemia associated with obesity and T2DM in humans.
The relationship between obesity and type 2 diabetes—the role of gut factors

Dietary interventions that alter the gut microbiota and host metabolism

Unlike the human genome, the composition of the intestinal microbiota can be modified and may depend on specific dietary habits. De Filippo et al. compared the gut microbiota in a group of children living in Burkina Faso, a country in West Africa (diet low in fat and animal protein; rich in starch, fiber, and plant polysaccharides; and predominantly vegetarian) and in a group living in Italy (diet high in animal protein, sugar, starch and fat, and low in fiber). Actinobacteria and Bacteroidetes were more represented in the microbiota of children from Burkina Faso than from Italy (10.1% vs. 6.7% and 57.7% vs. 22.4%, respectively), whereas Firmicutes and Proteobacteria were more abundant in the children from Italy than in children from...
In a representative animal model created by transplanting fresh or frozen adult human fecal microbial communities into germ-free mice, the structure of the microbiota was shifted within a single day when these mice were switched from a low-fat, plant polysaccharide-rich diet to a high-fat, high-sugar “Western” diet [58]. Moreover, individuals with a low bacterial gene count in the gut subjected to a 6-week energy restricted high-protein diet developed an increase in gene richness in the gut and subsequently led to a significant decrease in adiposity measures (hip circumference and total fat mass) and circulating cholesterol [59]. This plasticity of the gut microbiome may allow manipulation of certain gut microbiota associated with host disease, or conversely, enable enhancement of gut microbiota promoting health [36]. Figure 31.3 illustrates some of the potential beneficial effects of antibiotics, prebiotics, and probiotics on some metabolic disturbances related to obesity and T2DM.

Among the tools to modulate the gut microbiota, prebiotics (i.e., nondigestible food ingredients that benefit the host by selectively stimulating the growth and/or activity of one or a limited number of the bacteria in the gut) [60], and probiotics (i.e., live microorganisms which, when given orally in quantities adequate to allow colonization of the colon, confer a health benefit to the host) [61], are the most important. Gut microbiota modulations using prebiotics (i.e., short-chain inulin-type fructans, oligofructose, or wheat-derived arabinoxylan oligosaccharides) improve gut barrier functions, metabolic endotoxemia, and inflammation in obesity and T2DM [41].

Incretin hormones potentiate glucose-induced insulin secretion in response to a meal and are responsible for approximately half of the insulin response after food ingestion [35]. The gut microbiota can affect incretin function as demonstrated by an increased number of L cells in rats fed oligofructose (OFS) prebiotics, which promote increased levels of Bifidobacterium [62]. However, this relationship has not been established in humans and the role of the prebiotic on GLP-1 production independent of Bifidobacterium is unclear.

In mice, glucagon-like peptide-2 (GLP-2) antagonist treatment completely blocked the improvements in gut barrier function and the reductions in metabolic endotoxemia and inflammation induced by prebiotics, suggesting that the beneficial gut microbiota changes work through a mechanism involving GLP-2-driven improvement of gut barrier functions in obesity and T2DM [54]. Other preclinical studies have demonstrated that prebiotic-induced changes in the gut microbiota also modulate GLP-1, PYY, GIP, and ghrelin [41]. In human studies, the gut fermentation of nondigestible carbohydrates increases satiety after a meal and decreases food intake [63]. Parnell and Reimer demonstrated that gut microbiota modulations (after 12 weeks of prebiotic treatment) modulated ghrelin and PYY in overweight and obese patients and could thereby promote weight loss and improve glucose regulation [64].

Among the metabolites produced by the gut microbiota, the SCFAs could be a link between the gut microbiota and changes in gut peptides levels. Modifications in acetate, butyrate, propionate, and total SCFA levels occurred in the fecal samples of healthy humans treated with xylo-oligosaccharides alone or with an inulin/xylo-oligosaccharide mixture. Furthermore, an inulin/xylo-oligosaccharide mixture decreased plasma LPS levels and reduced the pro-inflammatory effects of LPS on cytokine gene expression in the blood of healthy humans [65]. Studies in mice demonstrated that administration of a probiotic promoted the release of the hormone GLP-1, resulting in reduced food intake and improved glucose tolerance [66]. More studies are needed to unravel the beneficial effects of prebiotics and probiotics in obese and type 2 diabetic patients.

---

**Figure 31.3** The modification of gut microbiota with antibiotics, the relative increase in Bifidobacterium by the use of prebiotics, and the use of probiotics such as Bifidobacterium and Lactobacillus shows beneficial effects on metabolic disturbances. Source: Esteve E, et al: *Current Opinion in Clinical Nutrition & Metabolic Care* 2011;14(5):483–490. Reproduced with permission of Lippincott Williams & Wilkins.
Summary

More and more evidence is emerging as to the importance of gut factors in the development of obesity and diabetes. Gut hormones have important roles in the regulation of appetite, food intake, and insulin secretion. Thus many of these hormones have been targets of pharmacotherapy for the treatment of obesity and diabetes. In particular, GLP-1 agonists have been an important addition to the armamentarium used to treat diabetes. Gut hormones are also becoming increasingly important as we begin to better understand the metabolic changes that occur after bariatric surgery. Moreover, gut microbiota have been implicated in a variety of human diseases including obesity and diabetes. They have a role in the defense against pathogens at the gut level, energy harvest, the development of the intestinal microvilli, maintenance of the intestinal barrier, and the synthesis of several vitamins. Currently, much of the data in humans remains controversial as to the cause and effect of the intestinal microbiota. However, the potential to alter microbiota as a means to improve health remains an important field of research.

References

CHAPTER 32

The role of energy metabolism in the regulation of energy balance

Charmaine S. Tam1,2 and Eric Ravussin1
1Pennington Biomedical Research Center, Louisiana State University, LA, USA
2The Charles Perkins Centre and School of Biological Sciences, University of Sydney, Sydney, NSW, Australia

Key points
- Total daily energy expenditure is comprised of resting metabolic rate (60–70%), diet-induced thermogenesis (10%) and structured and spontaneous activity thermogenesis (20–30%).
- Up to 85% of resting metabolic rate can be explained by fat-free mass, fat mass, age, and sex. Most of the remaining RMR variance is biologically (genetically) determined.
- Potential physiologic mechanisms leading to weight gain include low metabolic rate, low levels of physical activity, low sympathetic nervous system activity, and low fat oxidation, although more longitudinal data is still required.
- However, the major trigger of weight gain is excess food intake driven by the ubiquitous presence of cheap and palatable food.
- Given solid advances in gene targeting and related technologies, knockout and transgenic mouse models have identified molecular pathways that stimulate thermogenic programs. Further examination of these pathways including “browning” of white adipose tissue may prove useful for identifying new therapies for treating human obesity.
- Using analysis of covariance is the best way to normalize energy expenditure data for metabolic size and to compare differences between groups before and after an intervention.

Introduction

Body weight is an intricate balance between energy intake (calories consumed) and energy expenditure (calories burned). When energy intake exceeds energy expenditure over an extended period of time, body weight is gained and may lead to overweight and obesity. Whether the culprit of this weight gain is increased energy intake or reduced energy expenditure (EE) is unknown, but it is likely to be a combination of both with proportions of each varying from case to case. Since Lavoisier’s famous statement “la réspiration est donc une combustion” in 1790, generations of scientists have studied how the human body metabolizes energy and the most precise method to measure this metabolism. Progressing towards the development of the metabolic chamber in the 1970s and doubly labeled water in the 1980s, many research laboratories in the world can now precisely and accurately assess EE in laboratory and free-living conditions. This chapter will review the methods by which EE can be measured in humans and the physiologic and molecular mechanisms linking EE with excess weight gain. Lastly, given that metabolic carts are now widely available in research and clinical settings, an updated summary of the best way to analyze EE data in humans is provided.

Methods of measuring energy expenditure

Several methods have been developed to measure daily EE in humans. The most accurate methods involve continuous measurement of heat output (direct calorimetry) or gas exchange (indirect calorimetry) in individuals confined to small metabolic chambers. Confined individuals, however, are unable to continue activities related to daily living. As such, a number of field methods have been developed for this purpose ranging from the most accurate, doubly labeled water, to the less accurate, portable devices and self-reported questionnaires.

Direct calorimetry

Heat is the ultimate fate of all the body’s metabolic processes; EE can therefore be measured directly as heat loss using direct calorimetry [1]. Direct calorimetry involves putting individuals in a small insulated chamber where heat released in the form of dry heat (radiative and convective heat losses ~80%) or evaporative heat (heat loss from the evaporation of water from the skin...
and lungs (~20%) is measured. Direct calorimetric methods have largely become obsolete as they are expensive to build, run and maintain. Thus only specialized laboratories where direct heat loss values are required still have such capacity.

**Indirect calorimetry**

Under normal physiologic conditions, neither oxygen nor carbon dioxide are stored in the body. Therefore an indirect way of measuring EE is to measure oxygen consumption, carbon dioxide production, and nitrogen excretion. Presently, indirect calorimetry is widely used in normal and diseased conditions and is mainly performed using the open-circuit system which can measure EE over several hours or days in a respiratory chamber [1,2]. The open-circuit system can be performed using a mouthpiece, mask, transparent hood, or canopy. The individual inhales ambient air with a known composition. Changes in oxygen and carbon dioxide percentages in the expired air compared with the percentages in inspired ambient air reflect ongoing energy metabolism. During the last three decades, the indirect calorimetry method has been applied to confined rooms called respiratory or metabolic chambers [3]. These calorimeters are large enough (12,000–30,000 L) for an individual to live comfortably in for a few days, although typically most laboratories perform the measurement for 24 hours. Respiratory chambers can measure various components of daily EE including sleeping metabolic rate, the energy cost of arousal, the thermic effect of food, and the energy cost of spontaneous physical activity. With an average CV of 1–3% [4], respiratory chambers are used as the gold standard to validate newer methods of measuring EE such as doubly labeled water [5,6]. Metabolic chambers have become increasingly available in the past decade with more than 25 chambers worldwide in 2012 (there were only two worldwide in 1970).

**Doubly labeled water**

Developed by Lifson, the doubly labeled water (DLW) method is a form of “indirect” calorimetry and is based on the differential elimination of two nonradioactive isotopes, deuterium (2H) and 18 oxygen (18O) from body water following a single dose of these stable isotopes [7]. Oxygen tagged with the 18O tracer will equilibrate not only in body water but also in circulating bicarbonate and expired CO2. Over time, the 18O tracer in body water will decrease as CO2 is expired and water is lost in respiration, perspiration, and urine. The hydrogen molecule tagged with the 2H tracer will distribute only in the circulating water and bicarbonate and over time will decrease as water is lost. The isotopes are eliminated at different rates from the body, and thus the difference between the two elimination rates is a measure of CO2 production, from which total daily EE can be calculated using classical indirect calorimetry equations [6].

The major advantage of the DLW method is that it provides an integrated measure of total CO2 production over 4 to 21 days, yet requires only periodic sampling of blood, saliva, or most commonly, urine. Importantly, the DLW method allows individuals to be studied in the free-living state and has been validated repeatedly with excellent accuracy (1–3%) and precision (2–8%) against the gold standard indirect calorimetry [5,6]. The method is noninvasive and can be used in pregnant women [8], infants [9] and children, subjects with gastrointestinal disorders and the elderly. Limitations of DLW are the purchasing of expensive isotopes (2H), the measurement of analytical isotopic enrichments are labor intensive, and the inability to quantify substrate oxidation.

**Portable devices**

Pedometers and accelerometers are devices worn by an individual to quantify movement under free-living conditions. Pedometers assess displacement of the body with a single stride, with the output presented as steps taken or steps per day. Pedometers are economically feasible and therefore are easily applied in large populations. However, pedometers cannot discriminate between different activities or the intensity of physical activity and do not provide any quantification of EE per se.

Accelerometers assess and quantify the motion and movement associated with physical activity. Acceleration is the change in velocity with respect to time (m s−2), enabling accelerometers to quantify the intensity of movement. The majority of accelerometers are unilateral and sensitive to movement in the vertical plane (uniaxial accelerometer), but some are also sensitive to movements in the antero-posterior and/or lateral plane (triaxial accelerometer). Accelerometers provide a count value which describes the intensity, frequency, and duration of physical activity. However, the reliability of accelerometers to estimate daily EE varies from no correlation of activity counts to EE [10] to triaxial accelerometers explaining up to 90% of EE [11]. Such variability can be explained by several reasons including that validation experiments are usually performed in laboratory settings, with protocols consisting of treadmill-based locomotion whereas accelerometers are usually worn in free-living conditions. Also, it may be due to differences in monitors, duration (number of days), protocols (number and placement of monitors), methods of analysis and data interpretation [12]. It appears that pedometers and accelerometers provide useful indicators of physical activity levels but interpretation should be restricted to monitor-based output (e.g. steps taken, count value), as the validity of derived measures including total EE is often compromised [12].

**Questionnaires and activity logs**

Activity questionnaires, including interviews and diaries, are the most common tools for assessment of physical activity, with the energy cost of each activity estimated from energy expenditure tables. This method, called the factorial method, although cheap and easy to apply takes extended amounts of time for data analysis. Factorial assessments of a subject’s EE require that the type and duration of physical activities be recorded over a specified time period (usually several days). The energy cost of
each activity is estimated using energy-equivalent tables and multiplied by the time spent doing each activity. These values are then summed to derive an estimate of the EE related to the activity. As expected, this method is time-consuming for both participants and investigators and substantial errors can be incurred due to both inaccurate recall by individuals and the availability of numerous energy coefficients that can be applied to determine the energy cost of the activities. Comparisons with the doubly labeled water method generally show low correlations with systematic underestimates, overestimates, or agreement at the group level but large error at the individual level [13].

Components of energy expenditure and their relevance for human obesity

Total daily energy expenditure varies (TDEE) substantially in humans such that two adults with the same body size and body composition can have an EE that varies by as much as 1500 kcal d−1. The largest determinants of EE are weight, specifically fat-free mass, age, and sex [3]. Height and weight are all associated with higher TDEE, whereas TDEE declines with older age. Across all ages TDEE is approximately 11% higher in males, after adjusting for body size [14]. Understanding the inherent variability in EE will potentially assist in elucidating dynamics of excess weight gain and obesity. Such variability in daily energy requirements may be attributed to differences in the three components of EE: basal metabolic rate (BMR), the thermic effect of food and especially the energy cost of physical activity, consisting of both exercise and nonexercise activity thermogenesis (Figure 32.1).

Resting metabolic rate

Resting metabolic rate (RMR), accounting for 60–70% of TDEE, represents the energy required to maintain essential functions of the body’s vital organs and varies among individuals of different body size and body composition [3]. Under standard conditions, RMR is measured using the ventilated hood technique while the participant is awake and resting in a thermoneutral environment following an 8–12-hour fast and at least 12-hour abstinence from exercise and smoking. It has long been recognized that there is a close relationship between RMR and body size; this association has led to the development of widely used equations to predict RMR using height and weight from the classical Harris and Benedict equation proposed in 1919 [15], the Schofield equation in 1985 [16], and many others [17–19]. Up to three quarters of the variance in RMR is determined by fat-free mass [20,21] and to a lesser extent, by fat mass, sex, and age. Together, these four components explain up to 85% of the interindividual variance in RMR [20,21]. Eleven percent of the remaining variance can be further explained by family membership, therefore genetic factors also influence the variability in RMR [20]. It is also important to mention that fat-free mass is not metabolically uniform since it is composed of different tissue and organs, and thus tissue or organ-specific metabolic rates will contribute to further variability in measured RMR [22–24].

Diet-induced thermogenesis

Diet-induced thermogenesis, or the thermic effect of food (TEF) is the increase in EE associated with the digestion, absorption, and storage of food. It accounts for 5–15% of TDEE and is therefore the smallest component of daily EE. TEF is the most difficult and least reproducible component of EE to measure [4,25]. In contrast to RMR with a CV of 3–8% [26], the CV of TEF measured with indirect calorimetry is typically around 20% [25,27]. The measurement of TEF is performed in response to a test meal, either continuously or intermittently for 3–6 hours, under similar conditions to an RMR (fasting conditions at rest in a thermoneutral environment). Compared to a continuous measurement of TEF, intermittent measures of TEF have been shown to give a 50% lower response but a more replicable measure [27]. The duration of measurements
also contributes to variation in TEF, with shorter measurements (<5 h) yielding greater errors [28].

Given the variability in measuring TEF, its role in the etiology of obesity is still not clear. A review of 49 studies comparing TEF in lean vs. overweight/obese individuals reported that obesity was associated with an impaired TEF response, and was related to the degree of insulin resistance [29]. However, a subsequent review of 30 studies identified substantial shortcomings in the methods used to measure and calculate TEF, questioning the role of a low TEF in the development of obesity [30]. Indeed, the measurement of TEF is influenced by a myriad of factors including meal size and delivery (standard meal or glucose load), nutrient composition (protein exerts higher TEF), palatability of the meal, meal frequency as well as an individual's genetic background, age, physical activity, and sensitivity to insulin [29–32]. Importantly, prospective studies have not identified a relationship between TEF and weight change [25,33]. It is safe to say that a decrease in TEF equates to only a small amount of calories and therefore is very unlikely to explain significant degrees of obesity.

Activity thermogenesis
Activity thermogenesis, defined as body movement produced by skeletal muscle resulting in EE, comprises two distinct types of energy thermogenesis — the energy expended during exercise or structured physical activity (normally planned activities) and the energy expended in all other nonexercise activities (normally unplanned activities). The latter was described as spontaneous physical activity [34], and subsequently called Non Exercise Activity Thermogenesis (NEAT) by Levine and colleagues in 1999 [35]; it includes activities such as the energy cost of sitting, standing, walking, talking, fidgeting, and so on. Of course the hypothesis that reduced physical activity is the cause of the worldwide obesity epidemic is an attractive one. However, the amount of structured physical activity has remained relatively stable over the years and thus occupational physical activity has an even greater potential to impact energy balance [36]. Moreover, for the vast majority of people living in developed countries, exercise-related activity thermogenesis is negligible. With the exception for devoted exercisers, NEAT is the predominant component of activity thermogenesis. Not only that, NEAT is the most variable component of TDEE and varies by 200 kcal d−1 [37].

Perhaps some of the most carefully controlled studies examining the role of NEAT in weight gain have been performed by Levine et al. [35,38]. Levine found that resistance to the development of obesity may be caused by the ability of individuals to increase NEAT [35,38]. After 8 weeks of overfeeding (1000 kcal d−1), participants who were able to increase NEAT gained the least fat mass. In contrast, those who failed to increase NEAT in response to overfeeding gained the most fat mass. Moreover, changes in basal metabolic rate and TEF did not predict changes in fat gain [35]. In a follow-up study, Levine and colleagues attempted to identify whether there was a defect in NEAT in obese individuals. Using an integrated system of microsensors in undergarments to assess body posture and movements for 10 days, they found that obese individuals remained seated 2.5 hours longer than lean individuals, equivalent to an energy surplus of approximately 350 kcal d−1. In other words, if the obese subjects were to adopt the same sitting pattern as their lean counterparts, they may expend an extra 350 kcal each day. Next, they examined whether weight loss in the obese participants and weight gain in the lean would reverse this usual pattern of posture allocation. Similar to previous studies [39], they showed that both the lean and obese groups maintained their original posture allocation suggesting that some of the interindividual variation in NEAT can be explained by a genetic component [38]. Taken together, these studies suggest that promoting NEAT in our daily environment may be significant but is unlikely to be a real target to manage weight control. Furthermore studies assessing the effects of interventions on NEAT require careful and accurate measurement of TDEE (usually by DLW) as well as assessment of the other components of TDEE (RMR and volitional physical activity).

Energy expenditure in the etiology of obesity
Cross-sectional studies that compare lean and obese individuals have added little to our understanding of the role of EE in the progression of obesity [40]. Understanding the pathogenesis of human obesity demands longitudinal studies to reveal predictors or risk factors. The Pima Indians living in south-western Arizona are one of the most obese populations in the world with the highest reported prevalence of type 2 diabetes [41] and provide opportunities to examine predictors of weight gain. Risk factors related to EE include low metabolic rate, low activity thermogenesis, low sympathetic nervous system activity, and low fat oxidation [40].

Low metabolic rate
Obesity is associated with a high absolute metabolic rate, both in resting conditions and over 24 hours and therefore was not thought to be a potential factor involved in the pathogenesis of weight gain and eventually of obesity. Whether a low energy metabolism is involved or not in the propensity to weight gain is, however, still controversial [42]. One of the reasons for this controversy may be related to the way EE is compared in people of different body sizes. It is obvious that RMR is higher in obese individuals in absolute terms, but when adjusted for differences in body size (i.e., RMR divided by body weight, surface area, FFM, or even adjusted for body size by multiple regression analysis), the results are not always clear. Indeed, there is wide variability in the association between metabolic rate and body size, such that two individuals with the same fat-free mass and fat mass can differ in metabolic rate by up to 500 kcal d−1 [43].
In 126 adult nondiabetic Pima Indians, body composition and metabolic rate (ventilated hood) was measured at baseline and follow-up 4 years later [44]. Using an arbitrary definition of weight gain of 10 kg, subjects were divided retrospectively into gainers (n = 15) and nongainers (n = 111). Despite the gainers and nongainers having similar body composition at baseline, the “weight gainers” had metabolic rates which were 100 kcal d⁻¹ lower than the nongainers, suggesting that a low “relative” metabolic rate predicts weight gain. This was further supported by tertile analysis in the whole group, demonstrating that the risk of weight gain at follow-up was approximately seven times greater in subjects with the lowest metabolic rate compared to those with the highest metabolic rate [44]. In another group of 95 Pima Indians where 24-hour EE was measured in a metabolic chamber, 24hrEE adjusted for body size, age, and sex correlated negatively with the change in body weight and rate of change in body weight 2 years later [44], a relationship not found at all in other populations [45–47]. Taken together, these results demonstrate that low rates of EE (adjusted for body size, age, and sex), at least in populations more genetically susceptible to developing T2DM, are significant predictors of weight gain. Most interestingly, in response to weight gain, the mean adjusted metabolic rate of individuals who gained weight became similar to the mean adjusted metabolic rate of individuals who remained weight stable. This implies that weight gain may be a compensatory mechanism that results in an increased rate of EE engaging a mechanism of resistance to further weight gain (Figure 32.2). This “metabolic adaptation,” defined as a greater than predicted increase in EE beyond changes in fat-mass and fat-free mass, is highly variable between individuals and occurs in both lean and obese individuals. Similarly, a reduction in EE in response to weight loss can counteract further weight loss or even predispose to weight regain [48]. Importantly, a low metabolic rate explains a relatively low percentage of the variance in weight gain, and therefore other factors such as excessive energy intake and reduced activity levels are likely culprits [49].

**Low activity thermogenesis**

The energy cost of exercise and nonexercise activity thermogenesis accounts for 10–20% of TDEE. The most variable component of activity thermogenesis is NEAT, also known as spontaneous physical activity (SPA) [39]. Using whole-room calorimetry, longitudinal studies in the Pima Indians demonstrated that SPA, defined as the percentage of time that subjects were active in the chamber, accounted for a significant portion of 24hEE in males and females [39]. Fifty-seven percent of the variance in SPA was related to family membership, that is, probably to genetic background. Furthermore, SPA correlated inversely with the rate of weight change assessed 3 years later in males [39]. It can be argued that the confined environment of a respiratory chamber may not reflect physical activity levels in free-living conditions. However, physical activity measured in a respiratory chamber correlated with habitual physical activity level assessed by doubly labeled water in 50 nondiabetic Pima Indians [50]. Several longitudinal studies have clearly demonstrated the association between reduced NEAT/SPA and weight gain [38,50]. Whether a similar association exists for habitual or structured physical activity is not clear [51–53]. A study in 92 nondiabetic Pima Indians showed that physical activity level (assessed by doubly labeled water) was not associated with changes in body weight 4 years later [49].

**Low sympathetic nervous system activity**

Reduced sympathetic nervous system activity (SNS) plays a causative role in several rodent models of obesity [54]. In humans, SNS activity can be directly measured in the muscle by microneurographic recordings [55]. This muscle sympathetic nervous activity (MSNA) correlates well with indirect measurements of SNS activity such as plasma norepinephrine turnover and plasma norepinephrine activity. In Caucasians, SNS activity is independently related to total daily EE [55,56] and to each of the major components of (TDEE-RMR) [55], TEF [57] and SPA [58]. However, such associations were not seen in Pima Indians [56], probably because Pima Indians have lower SNS activity compared to Caucasians [55]. Given that Pima Indians also have lower rates of total daily EE preceding weight gain [44] as well as reduced sympathetic nervous system activity, these two factors may represent potential mechanisms predisposing this population to body weight gain [55]. Indeed, baseline SNS activity (24-hour urinary epinephrine excretion rates) was negatively associated with weight gain after a 3-year follow-up [59]. Because the latter observation has not been
reported in other populations, the case for whether a reduced sympathetic nervous system activity precedes weight gain is still unclear.

**Low fat oxidation**

The choice of nutrient substrate (carbohydrate vs. fat) used for oxidation may be an important factor in the etiology of obesity and is calculated by the ratio between carbon dioxide production and oxygen consumption, the “nonprotein respiratory quotient (RQ).” RQ reflects the ratio of carbohydrate to fat oxidation and can vary between ~0.7 after an overnight fast when fat is the primary substrate to values close to 1.0 after a high-carbohydrate meal when glucose is the principal energy substrate. Apart from the obvious influence from diet composition, other factors that may influence RQ are recent energy balance, sex, adiposity, and family membership, suggesting a genetic component to RQ [60]. Low fat oxidation has been suggested to be a factor in the development of obesity. In 111 Pima Indians, 24-h RQ (measured in a metabolic chamber) was correlated with subsequent changes in body weight and fat mass at follow-up 3 years later. In fact, subjects in the top 90th percentile of RQ ("low fat oxidizers") had 2.5 times greater risk of gaining 5 or more kilograms compared to individuals in the bottom 10th percentile of RQ ("high fat oxidizers") [60]. Similar results have also been reported in Caucasians [61].

**Molecular mechanisms of energy expenditure variability and avenues for treatment**

**Leptin**

Leptin, an hormone that is predominantly secreted by adipocytes and primarily binds to receptors in the hypothalamus as well as in the entire central nervous system, plays a key role in regulating long-term energy homeostasis in rodents and humans. Humans with the ob mutation lack circulating levels of leptin and therefore develop severe, early-onset obesity with associated metabolic and behavioral consequences including hyperphagia, defective thermogenesis, and T2DM. Leptin replacement therapy using daily subcutaneous injections completely reverses these symptoms in patients with congenital leptin deficiency as well as lipodystrophy, largely through the dramatic weight loss that occurs in these patients [62, 63]. These early findings in leptin-deficient patients suggested that administering leptin in obese patients would be a stand-alone magic bullet for the treatment of obesity. However, disappointingly this was short-lived. In 1999, a randomized control trial of daily subcutaneous recombinant leptin injection was performed in 54 lean and 73 obese subjects. In the initial phase of the study lasting 4 weeks, lean and obese subjects lost similar amounts of weight with leptin [64]. Obese subjects were studied for a further 20 weeks. Of the 47 patients who completed the study, the eight receiving the highest dose of leptin lost 7.1 kg while the placebo lost 1.3 kg. However, this weight change varied widely among patients, from a loss of about 15 kg to a gain of 5 kg in the group treated with the highest dose. Moreover, these doses induced skin irritation and swelling at the injection site in 62% of patients and headache in half the patients. The potential of leptin therapy as a panacea for obesity appeared to be an early failure. However, leptin replacement therapy may be more pertinent for patients during weight loss maintenance. In calorie-restricted animals and humans, exogenous leptin administration has been shown to reverse the metabolic adaptation induced by caloric restriction and restore EE, catecholamine and thyroid hormone levels, and skeletal muscle efficiency back to baseline values [65, 66].

**β3-adrenergic receptors**

β3-adrenergic receptor (β3-AR) agonists are very effective thermogenic as anti-obesity and insulin-sensitizing agents in rodents with the main sites of action in skeletal muscle as well as white and brown adipocytes. Testing in humans with the first generation of β3-AR agonists, such as BRL 26830A and CL 316243, revealed promising results on energy metabolism. However, these compounds were not selective enough for the β3 receptor and shared substantial affinity for the β1- and β2-AR subtypes thus inducing many undesired effects such as accelerated heart rate, tremors, and hypokalemia [67]. The second generation of β3-ARs had improved selectivity but poor oral availability or pharmokinetics. Given these problems, we are not aware of any β3-ARs that have progressed beyond Phase II clinical trials [68]. However, given the recent findings that brown adipose tissue (BAT) is present in adult humans [69–71], the potential of β3-ARs agonists for increasing the amount and activity of BAT is an active area of exploration. Furthermore, a novel trend is now to target a subpopulation of adipocytes in the white adipose tissue to become more like brown adipocytes and therefore oxidize more fat.

**Uncoupling proteins**

The uncoupling proteins (UCPs) gene and especially UCP1 encode mitochondrial protein carriers, which uncouples respiration from ATP production stimulating heat production. Activation of such proteins may therefore be important for the control of metabolic efficiency. Discovered in 1978, UCP1, is only expressed in BAT and plays an important role in thermogenesis in rodents [72]. UCP1 was generally believed not to play a major role in energy balance as adults were thought to have very little BAT stores. However recent studies using 18F-FDG labelled glucose in PET/CT revealed the presence of BAT in adult humans, the prevalence of which is estimated to be between 2 and 100% [69–71]. Currently, there is intense investigation into mechanisms for stimulating existing BAT or to induce BAT in white adipose tissue as a means to promote thermogenesis and EE [73].

In the late 1990s, two new mitochondrial uncoupling proteins were discovered, UCP2 (widely distributed in a variety of tissues) and UCP3 (abundantly expressed in skeletal muscle) stimulating great excitement into these new UCPs to explain
some of the variability in energy metabolism in humans. Transgenic mice with very high overexpression of UCP3 in their skeletal muscle were resistant to diet-induced obesity and had improved insulin sensitivity compared to their wild-type counterparts [74]. In humans, UCP2 and UCP3 expression levels were measured in skeletal muscle of 19 non-diabetic male Pima Indians covering a wide range of body weight. BMI was negatively associated with UCP3 expression, and metabolic rate during sleep, adjusted for fat-free mass and fat mass, was positively correlated with UCP3; no associations were observed between UCP2 and BMI or metabolic rate [75]. Whether UCPs play a significant role in the etiology of human obesity remains to be established.

**Diacylglycerol transferase and acetyl CoA carboxylase**

Triglyceride synthesis has been implicated to occur though the acyl CoA:diacylglycerol transferases (DGATs), enzymes that catalyze the final and only committed step in the glycerol phosphate pathway. DGAT1−/− mice are leaner than wild-type mice and have smaller adipocytes. When fed a high-fat diet, DGAT−/− mice are resistant to obesity and are protected from diet-induced hepatic steatosis likely due to increased EE. DGAT1−/− mice have increased spontaneous physical activity, increased expression of UCP1, and increased leptin sensitivity [76, 77]. Conversely, overexpression of DGAT1 in adipose tissue results in impaired insulin sensitivity and liver steatosis [78]. These findings in rodent models where DGAT has been deleted suggest that pharmacologic inhibition of DGAT may be a novel therapeutic target for obesity in humans. Preclinical studies using DGAT inhibitors, such as T863, inhibit lipid absorption, causes weight loss, improves insulin resistance, and alleviates hepatic steatosis in high-fat-diet fed mice [79]. However, we are not aware of DGAT inhibitors in human clinical trials.

**Acetyl CoA carboxylase**

Acetyl CoA carboxylase (ACC), which exists in at least two isoforms in humans, is responsible for synthesizing malonyl coenzyme A, a potent inhibitor of fatty acid oxidation. ACC1 is expressed mainly in liver and adipose tissue, and ACC2 in the heart and skeletal muscle. Although mice lacking ACC1 die young, ACC2−/− mice have increased fat oxidation and therefore decreased weight and fat stores compared to wild-type mice. Furthermore, when exposed to a high-fat diet, ACC2−/− mice are resistant to obesity and the development of diabetes [80–84]. This preclinical evidence suggests that ACC inhibitors may serve as therapeutic targets for obesity and clinical trials are ongoing.

**Analysis of energy expenditure data in humans**

Regardless of the technique used to measure daily EE, adjustment for metabolic body size is essential for valid interpretation of the data. Traditionally, investigators have used the ratio method and simply divided EE by the amount of metabolic mass (body weight, fat-free mass, or surface area. However, this leads to erroneous results [85–87] since this method fails to take into account the nonzero intercept in the relationship between EE and metabolic mass. As a consequence, the adjusted metabolic rate is artificially increased as the fat-free mass decreases. An alternative and more valid approach is the regression-based approach which derives prediction equations from simple available measures (sex, age, FM, and FFM) to determine size-adjusted EE [85–87]. As one would expect, the method chosen for normalization of metabolic size is particularly crucial when assessing the effects of an intervention (weight loss, drug intervention, etc.) on EE.

**Conclusions**

The global obesity epidemic has stimulated intense interest in the genetic and molecular basis of body weight regulation. Changes in the components of energy expenditure, comprised of basal metabolic rate, thermic effect of food, and physical activity thermogenesis have important roles in energy balance. However, what ultimately determines “metabolic efficiency” is extremely complex. All components of energy expenditure are likely influenced by interactions between our individual genes and the environment affecting many physiological and biochemical processes. Recent progress in gene targeting and other technologies has pushed mouse models to the forefront of this effort. Whether these preclinical findings will translate into therapeutic utility for human obesity is not yet known. There is no doubt that the field of energy expenditure has developed tremendously over the years and understanding the homeostatic balance between energy intake and energy expenditure will continue to be an active research endeavor.

**References**


50 Snitker S, Tataranni PA, Ravussin E: Spontaneous physical activity in a respiratory chamber is correlated to habitual physical activity. *International Journal of Obesity and Related Metabolic Disorders* 2001;25(10):1481 – 1486.


### Key points

- Diet is the cornerstone in obesity management, and compliance to a reduced energy intake the challenge.
- Low-energy diets and very-low-energy diets using meal replacements are effective to induce weight loss by ensuring a high degree of adherence.
- A greater initial weight loss is associated with greater long-term retention and weight maintenance.
- Different diets take advantage of an enhanced satiety effect by higher protein content, and lower content of fat, refined carbohydrates, and sugar-rich soft drinks.
- Higher protein, low-carbohydrate diets are the most efficient and generally produce beneficial effects on all cardiometabolic risk factors.
- Weight loss in type 2 diabetics can be induced by diet and exercise. It has beneficial effects on glycemic control and reduces the use of antidiabetic medications.
- Weight loss in type 2 diabetics also improves QoL, sleep apnea, and cardiometabolic risk factors, but does not reduce cardiovascular events and mortality.
- Some antidiabetic medications have weight-neutral and weight-lowering properties.
- A number of weight loss medications have favorable effects on weight loss also in diabetics, and improve glycemic control and cardiometabolic risk factors.
- A number of new pharmaceutical compounds that target both obesity and type 2 diabetes are under development and show promise.

### Principles of diet-induced weight loss

Dietary therapy consists of instructing patients on how to modify their dietary intake to achieve a decrease in energy intake whilst maintaining a nutritionally adequate diet and causing the fewest possible adverse effects on hunger and satiety. Obese patients have, due to their enlarged body size, higher energy requirements for a given level of physical activity than normal weight counterparts (Figure 33.1). Obese diabetics have slightly higher energy requirements than simple obese for a given body size and composition. Reducing the obese patient's total energy intake to that of a normal-weight individual will inevitably cause weight loss, consisting of about 75% fat and 25% lean tissue, until weight normalization occurs at new energy equilibrium. For patients with class I obesity this requires an energy deficit of 300–500 kcal d<sup>−1</sup>, and for patients with class III obesity 500–1000 kcal d<sup>−1</sup> until a satisfactory weight loss is achieved or an alternative approach is required.

---

**CHAPTER 33**

**Treatment of obesity: lifestyle and pharmacotherapy**

**Arne Astrup**

Department of Nutrition, Exercise and Sports, University of Copenhagen, Denmark

---

The more rapid the weight loss. A deficit of 300–500 kcal of energy intake. The larger the daily deficit in energy balance and energy requirements assessed by measurement of energy expenditure or by apparent energy intake during weight stability. The growing underreporting with increasing body fatness makes the self-reported energy intake invalid for estimation of energy requirements in obese patients.

There is no evidence that diet composition exerts clinically important effects on energy absorption and energy expenditure, but diet composition may affect hunger and satiety to such an extent that it facilitates reduction of total energy intake. Energy deficit can be achieved by setting an upper limit for energy intake. The larger the daily deficit in energy balance the more rapid the weight loss. A deficit of 300–500 kcal d\(^{-1}\) produces a weight loss of 300–500 g per week, and a deficit of 500–1000 kcal d\(^{-1}\) produces a weight loss of 500–1000 g per week. Greater initial energy deficits may produce even larger weight loss rates. Total energy expenditure declines and normalizes along with weight loss, and total energy intake must therefore be gradually further reduced to maintain the energy deficit. An alternative approach is to take advantage of the differences in the satiating power of the various dietary components in order to cause a spontaneous reduction in energy intake. This is the principle of ad libitum low-fat, low-carb, high-protein diets.

**Choosing the dietary energy deficit**

Initially the target of a weight loss program should be to decrease body weight by 10%. Once this is achieved a new target can be set. Patients will generally want to lose more weight and there is no evidence that unrealistically high weight loss expectations impair weight loss success [1], but it should be noted that even a 5% weight reduction improves risk factors and risk of comorbidities. However, several factors should be taken into consideration, including the patient’s degree of obesity, previous weight loss attempts, medications, risk factors, comorbidities, and personal and social capacity to undertake the necessary lifestyle changes.

To prescribe a diet with a defined energy deficit it is necessary to estimate the patient’s actual energy requirements. As assessment of the patient’s habitual energy intake from self-reported diet registration is invalid as obese individuals systematically underreport energy intake by up to 30–40%. Energy requirements should therefore be assessed indirectly by estimation of total energy expenditure. Resting metabolic rate (RMR) can be measured by indirect calorimetry, or be estimated with great accuracy using equations based on body weight, gender, and age [2], or even better, estimated from the size of fat-free mass and fat mass. Total energy expenditure (=energy requirement) is estimated by multiplication of RMR (kcal d\(^{-1}\)) by an activity factor (PAL; physical activity level) [2]. The energy level of the prescribed diet is defined as the patient’s energy requirement minus the prescribed daily energy deficit.

**Theoretical versus clinical outcome**

Translating physiologically based considerations regarding energy balance and weight loss into successful clinical practice requires a high degree of compliance on the part of the patient, which is difficult to obtain. Weight loss results tend to be much better in clinical trials conducted in specialized clinics than in trials conducted by nonspecialists without sufficient resources or access to auxiliary therapists (dieticians, psychologists, etc.). Compliance and adherence to the diet are the cornerstones of successful weight loss, and are the most complicated part of dietary treatment of obesity. To improve adherence the prescribed diet should match the patient’s food preferences and food culture, and take personal, educational, and social factors into account. The patient must be seen frequently and regularly.

Long-term weight reduction is unlikely to succeed unless the patient acquires new eating and physical activity habits. These behavioral changes should be an integral part of the treatment program (see Chapter 41 on Behavioral Therapy).

**Prognostic markers of weight loss success and reasons for failure**

There is considerable variation in the weight loss and weight maintenance achieved by patients enrolled in dietary treatment programs. Some variation can be attributed to physiological differences (genes, resting metabolic rate, age, gender) and some to differences in adherence (genes related to adherence, behavior, acceptability of diet), although combinations of and interactions between physiology and behavior may be the major factor. This variation is not apparent from the mean weight loss of a group, which is difficult to translate into a clinically relevant success rate based on the intention to treat principle, that is, how large a proportion of the patients entering a treatment achieve a certain weight loss, typically >5% and >10% of pretreatment body weight. It is difficult to identify clinically the patients who will benefit most before treatment is initiated. The most powerful predictor of long-term weight loss success is initial weight loss, that is weight achieved within the first 1–2 months of treatment (Figure 33.2 (b)).
Options for weight loss diets

Therapeutic weight loss diets distinguish between several recognized regimens. Low-energy diets (LED) usually provide 800–1500 kcal d−1, and use fat and carbohydrate-reduced foods. Very-low-energy diets (VLED) are modified fats providing 200–800 kcal d−1 that replace normal foods. Ad libitum low-fat diets do not restrict energy intake directly, but target a restriction of ad libitum fat intake to 20–30% of total energy intake. Energy intake is spontaneously reduced because of the higher satiating effect of this diet if protein intake is increased, and a modest weight loss occurs.

Very-low-energy diets (VLED)

Starvation (less than 200 kcal d−1) is the ultimate dietary treatment of obesity, but it is no longer used because of the numerous and serious medical complications associated with prolonged starvation. Starvation has been replaced by VLED (200–800 kcal d−1), which aims to supply very little energy but all essential nutrients. Reducing the energy content of a diet requires an increased nutrient density. This can be difficult to obtain with natural foods if the diet is to be acceptable once the energy content is less than 800 kcal d−1. This has led to the commercial production of VLED, supplemented with all nutrients in RDA amounts. For decades 250–400 kcal d−1 formula diets were extremely popular. The first VLED were clearly nutritionally insufficient, but reports of adverse effects and results from research have brought about a gradual increase in energy level. Today the 800 kcal d−1 VLED is the only version recognized as being both effective and safe. A number of studies have shown that VLED with energy levels of less than 800 kcal d−1 do not produce a greater weight loss and are less well accepted than those comprising 800 kcal d−1. VLED usually provide a ketogenic diet with an energy content of around 800 kcal d−1 in the form of nutrition powders, or in the form of protein, mineral, trace element, and vitamin-enriched meals or drinks. VLED can induce very rapid weight loss over a 2–3 month period (Figure 33.2(a)), but should only be used as an induction of weight loss, and should be followed by a weight maintenance program. This is not because a rapid initial weight loss causes poorer long-term weight maintenance per se, as initial weight losses are positively associated with long-term weight loss [1]. Obese patients having the greatest initial weight loss during an 8-week 800 kcal d−1 diet also retain the greatest weight loss during a subsequent maintenance program (Figure 33.2(b)).

VLEDs, compared with LEDs, induce significantly greater short-term weight losses, but similar long-term losses, and attrition has been found to be similar with VLED and LED regimens [4]. Some concern has been raised about the cardiac safety of the use of VLED with less than ∼600 kcal d−1 and patients using VLED have an increased risk of developing gallstones (Table 33.1). Their use without medical supervision has generally been abandoned and should not be recommended.

Low-energy diets (LED)

LED usually provide 800–1500 kcal d−1 and normally consist of natural foods. LED are also called “conventional diets” or “calorie-counting diets” (because previously more emphasis was put on restricting the total energy level of the diet and less on the macronutrient composition). Although macronutrient composition of the diet is of less importance for short-term weight loss it is now usually modified in order to maximize the beneficial effect on cardiovascular risk factors and insulin resistance, and to prevent cancers, and perhaps also promote long-term weight maintenance. For practical reasons LED are low-fat, protein-rich diets with a fixed energy allowance. An example of an appropriate nutrient composition for an LED is given in Table 33.2. LED should be supplemented with a daily vitamin and mineral tablet. LED for women is 800–1200 kcal d−1 and for men it is 1000–1500 kcal d−1. LED produce a lower rate of weight loss than VLED, but randomized clinical trials (RCTs) demonstrate that the long-term (>1 year) weight loss is no different from that of the VLED [1]. Furthermore, using LED introduces healthy eating habits early in the weight reduction program, giving a longer period in which to familiarize the patient with the dietary changes that are a central element in a weight maintenance program.

Based on 34 RCTs examining the impact of LED in obese subjects it was concluded that LED produces weight loss regardless of duration of treatment, and body weight was reduced by an average of 8% over 3–12 months. In trials lasting 3–6 months, and 6–12 months, LED produced mean weight losses of approximately 8% compared to controls [5]. Though they targeted the same energy deficit there was a large variation in mean weight loss between studies. The variation is due less to differences in patient selection than to differences in the skill and efforts of the therapeutic teams. The same large variation exists within each cohort of patients and the mean weight loss of 8% corresponds to approximately 80% of the patients losing >5% and 15% losing >10%. It is difficult to separate the impact of the dietary therapy because all trials included some elements of behavior modification, which clearly contributed to the results.

Ad libitum lower fat diets (25–30% energy from fat)

Efficacy

Data from animal and experimental research, observational studies and numerous RCTs, have shown that a high dietary fat content plays some role in the development of obesity [6–8]. A high-fat diet especially promotes weight gain and obesity in sedentary individuals with little self-restraint. The main mechanisms are passive overconsumption of energy promoted by the high energy density of fatty foods, and a reduced fat oxidation capacity in susceptible individuals.

The lower fat diet (25–30% of energy) is clinically relevant for obese patients who habitually consume a high-fat diet (35–50% of energy), but for the majority of the obese population only a
modest reduction in dietary fat content is possible and more emphasis should be put on nonfat dietary components, such as reduction of soft drinks, and modification of protein, glycemic index, fiber, and whole grain, and on fat quality, in order to ensure the lowest cardiovascular risk profile.

Lower-fat diets can be more satiating provided the fat is replaced by protein and to some extent by low-glycemic index carbohydrates, high fiber foods, and whole grains. Low-fat diets where fat is replaced by refined starch and simple carbohydrates will produce little or no weight loss, and have adverse effects on cardiovascular risk factors.

A meta-analysis of 34 ad libitum, low-fat interventions lasting >2 months and 35 control groups, found a statistically highly significant weight loss difference of 3.3 kg [9].

**Safety**

A change from a normal-fat to a low-fat, high-carbohydrate diet has been shown to induce hyperinsulinemia, hypertriglyceridemia, and low HDL-cholesterol concentrations, but these effects are highly dependent on the composition of the nonfat part of the diet [10,11]. However, although healthier high-fat diets, that is, replacement of saturated and trans fats with...
unsaturated oils instead of with carbohydrates, have proved to have beneficial effects on cardiovascular and diabetes risk profile, the higher energy density may increase risk of weight regain. Another example is the DASH-diet, a fat-reduced diet with increased intake of dairy products, fruit, and vegetables, that has been shown to produce a clinically relevant reduction in blood pressure, and other cardiovascular risk factors [12].

**Low-carbohydrate, high-protein diets (HPLCDs)**

Classification of high-protein, low-carb diets

There are a wide variety of popular diets and scientifically tested diets that are named either low-carb or high-protein diets (Atkin’s, Dukan, The Zone, South Beach, I-diet, paleolithic diet), but they share the common feature that they are prescribed “ad libitum,” that is, without energy restriction, and they work by suppression of hunger and enhanced satiety effect. They prescribe higher protein content either directly or indirectly through a marked restriction in total carbohydrate content, which automatically leads to higher protein content. They also stress carbohydrate quality, mostly by elimination of simple sugars, and either advocate carbohydrates with low glycemic index, high fiber and/or whole grain contents, and may achieve this by banning certain food groups that provide starchy carbohydrates (e.g. paleolithic diet). There is no official classification but the increasing common definition is that a low-carbohydrate diet has $\leq 45\%$ of energy from carbohydrates, low-fat diets have $\leq 30\%$ of energy from fat, and high-protein diets have $>20\%$ of energy from protein. However, the protein content of high-protein diets can vary from 20–50% of energy, and there are numerous combinations of low-carb, high-protein diets that can be either low or high fat. A general scientific assessment is therefore subject to reservations and the imprecise classifications give rise to inconsistent outcomes of meta-analyses of efficacy and safety. It is also necessary to distinguish between diets consumed ad libitum and those with a fixed prescribed energy intake.

**Efficacy of high-protein, lower carb diets**

A large number of RCTs have been conducted to compare HPLCDs with low-fat diets, and a number of systematic reviews and meta-analyses have assessed efficacy and short-term safety. Approaches have varied slightly, but generally these analyses have found that HPLCDs have a more favorable effect on weight loss, body composition, resting metabolic rate, and cardiovascular risk factors than fat-reduced diets.

Santos et al. conducted a meta-analysis of 23 RCTs, involving 1141 obese nondiabetic subjects, on “low-carb” diets, based on a selection of publications that named the diet “low-carb” no matter the composition and degree of carbohydrate restriction [13]. They found that HPLCDs produced a 7.0 kg greater weight loss than low-fat diets over 6–12 months, and that the difference was maintained for up to 2 years. HPLCD caused greater reduction in waist circumference (−5.74 cm), systolic blood pressure (−4.81 mmHg), diastolic blood pressure (−3.10 mmHg), plasma triglycerides (−29.71 mg dL$^{-1}$), fasting plasma glucose (−1.05 mg dL$^{-1}$), glycated hemoglobin (−0.21%), plasma insulin, and in plasma hs-CRP, and an increase in HDL-cholesterol. No differences were found in LDL-cholesterol and serum creatinine levels.

Krieger et al. used meta-regression of RCTs to determine the effect of changes in dietary protein and carbohydrate contents during energy restriction. They included a total of 87 studies, with 165 intervention groups, with diets providing at least 4200 kJ d$^{-1}$ (1000 kcal d$^{-1}$). Diets that provided less that 35–41.4% energy from carbohydrate were associated with a 1.7 kg greater weight loss, a 0.7 kg greater loss of fat-free mass, and a 2.0 kg greater loss of fat mass than were diets with a higher percentage of energy from carbohydrate. In studies lasting more than 12 weeks the effects were increased to 6.6 kg weight loss and 5.6 kg greater fat loss. Protein intakes of more than 1.05 g kg$^{-1}$ were associated with 0.60 kg additional fat-free mass retention compared to diets with protein intake of less than 1.05 g kg$^{-1}$. In studies with duration of more than 12 weeks this difference increased to 1.2 kg. Krieger et al. concluded that HPLCDs favorably affect body mass and composition independent of energy intake, which in part supports the proposed metabolic advantage of these diets [14].
Wycherley et al. looked at the effects of isocaloric, high-protein versus normal protein diets with matched fat contents (normal to low) [15]. Twenty-four trials included 1063 individuals. Higher protein diets produced more favorable changes in body weight (0.8 kg), fat mass (0.9 kg), and triglycerides than normal protein diets, and mitigated reductions in fat-free mass and resting metabolic rate. Changes in fasting plasma glucose, fasting insulin, blood pressure, and LDL, HDL and total cholesterol, did not differ between diet types.

Finally, Hu et al. compared the effects of low-carbohydrate diets (≤45% of energy from carbohydrates) with those of low-fat diets (≤30% of energy from fat) in a meta-analysis of RCTs [16]. Twenty-three trials with a total of 2788 participants were analyzed. Both diets lowered body weight similarly and improved cardiometabolic risk factors. However, the low-carbohydrate diets produced a lower reduction in total cholesterol and LDL-cholesterol, a greater increase in HDL-cholesterol, and a greater decrease in triglycerides.

Mechanisms of action of high-protein, low-carb diets

From a physiological perspective there is evidence that any reduction of energy intake, whether it is achieved by prescribed energy reduction or by enhanced satiety on an ad libitum diet, should ensure a minimum protein supply in order to avoid excessive loss of lean body mass and subsequent reduction in metabolic rate. It is quite clear that the current RDA for protein is insufficient to avert adverse effects on body composition during weight loss, and that 2 × RDA, that is, 1.6 g protein per kg BW (~ at least 20% of energy from protein) is required [17]. A higher protein content also enhances satiety compared to carbohydrate, and some of the effects can be attributed to a stronger stimulation of gastrointestinal satiety hormones (GLP–1, PYY) [18]. Realistic protein contents are 20–35% of energy. A pronounced reduction of carbohydrate content enhances weight loss and improves cardiometabolic risk, but such a diet is difficult for many obese patients to adhere to in the longer term. A more realistic diet allows more carbohydrate, provided this is in the form of low-glycemic index carbohydrates. The weight-reducing effects of increasing the protein and of lowering the glycemic index of the diet are additive [19], and the low GI component enhances glycemic control [20] and decreases inflammation [21]. Moreover, this diet composition is more commensurate with Western food culture and is highly accepted among obese patients [22].

In conclusion, HPLCDs are more effective than low-fat diets at reducing weight and improving metabolic risk factors. Low-carbohydrate diets could be recommended to obese persons with abnormal metabolic risk factors for the purpose of weight loss.

Safety of high-protein, lower carb diets

If protein is increased in the diet from standard recommended levels of 10–15% of total energy to 25–30%, and fat is held constant at 30% of total energy, the percentage of carbohydrate consumed is reduced to 40–45% of total energy. The shift in the proportion of protein and carbohydrate in the diet should be achieved by replacing soft drinks containing refined sugar with high-protein beverages like milk, and adjusting the proportion of carbohydrate (i.e. smaller portions of white bread, rice, and pasta) and protein (i.e. larger portions of lean meat, fish, dairy products, beans, and lentils) during meals. These dietary changes will lead to a spontaneous reduction in total energy intake and body weight, and maintain the nutritional quality of the diet.

Unless lean protein sources are consumed, and intake of unrefined carbohydrate is maintained, high protein intake may result in reduced consumption of fiber, vitamins, and minerals (i.e. potassium, calcium, and magnesium), which could negatively impact cardiovascular, bone, and kidney function. Although these are reasonable possibilities, particularly for the Atkin’s type diets with very low carbohydrate intakes, many studies suggest that replacement of some carbohydrate by protein has a neutral or even positive influence on inflammation, risk factors for T2D, bone health, and renal function [13,21,23–26].

Any assessment of HPLCDs’ effects on cardiovascular risk factors needs to address the weight loss produced. Weight loss has such a powerful, beneficial effect on most risk factors, that the weight loss produced by HPLCDs can easily mask any adverse effect of the diet composition on risk factors. The effects on risk factors should therefore be studied when weight loss has plateaued.

RCTs of up to 2 years duration have not revealed any adverse effects on cardiometabolic risks. Contrastingly, a number of observational studies have found associations between HPLCDs and cardiovascular risks [13]. However, in these longitudinal studies relying on self-reported food intake, very few participants consumed a HPLCD and the prediction of risk was based mainly on statistical extrapolations from subjects consuming high amounts of protein, saturated fat, refined carbohydrates, and so on. These studies also found that consumption of HPLCDs is associated with weight gain, but it is difficult to see how these findings can be causally related in light of the lack of plausible, physiologic mechanisms.

Kidney and bone health

It has been suggested that very high protein intakes in animal studies may damage the healthy kidney, reduce GFR, and produce albuminuria. The few RCTs that have investigated this question in humans have not found any detrimental effect of high-protein diets, but rather the opposite [23–25,27], probably due to the fact that diabetes, hypertension, and obesity are risk factors of loss of kidney function, and as long as HPLCs produce weight loss all risk factors for impaired kidney function improve.

HPLCDs are beneficial for bone health in that they provide adequate protein, calcium, and vitamin D, during weight loss and maintenance [26,28].
Dietary weight maintenance programs

In professional weight loss programs LED induce a 5% weight loss in almost all patients, and frequent clinical encounters during the initial 6 months of weight reduction appear to facilitate achievement of therapy goals. More ambitious success criteria (>10% weight loss) can be met by the majority of patients if the treatment program also includes group therapy and behavior modification. A number of studies suggest that the use of IT-technology can further improve the cost-benefit of the programs. The real challenge is to maintain the reduced body weight and prevent subsequent relapse (Figure 33.3). In a systematic review of long-term (>3 years follow-up) efficacy of dietary treatment of obesity, success was defined as maintenance of all weight initially lost or maintenance of at least 9 kg of initial weight loss [29]. Initial weight loss was 4–28 kg and 15% of the followed-up patients fulfilled one of the success criteria, and the success rate was stable for up to 14 years of observation. Diet combined with group therapy led to better long-term success rates (27%) than did diet alone (15%), or diet combined with behavior modification and active follow-up, although active follow-up produced better weight maintenance than passive follow-up (19% vs. 10%).

Whereas energy restriction (LED) is successful for weight loss induction independent of dietary composition, the medium-fat, high-protein/low GI diet seems to be more effective for long-term weight maintenance and prevention of weight regain. Weight maintenance diets need to be realistic, unrestrictive, and to fit into a food culture in order to be accepted over the long term and achieve compliance from the reduced obese subjects.

In the Diogenes European multicenter trial the impact of higher protein and lower glycemic index of diets for prevention of weight regain was studied in 800 obese families. After an initial weight loss of 11 kg achieved by an 800 kcal.d^−1 diet over 8 weeks in the obese adults they were randomized to five diets differing in protein content and glycemic index. After 8 months those randomized to higher protein and lower GI remained essentially weight stable, whereas those on diets with normal protein and higher GI regained weight [19]. The effect of higher protein (23% of energy) and lower GI was additive. This diet composition has also been shown to be effective in children [30], and to improve cardiometabolic risk [21]. The challenge is to incorporate the diet into a food culture in order to achieve long-term adherence.

Weight management programs for the prevention of type 2 diabetes

The potential of intensive lifestyle treatment to induce and maintain weight loss has been evaluated in two major studies on the prevention of T2D in obese subjects with IGT [31,32]. In the Finnish Diabetes Prevention study 522 overweight individuals with IGT were randomized to an intensive lifestyle intervention or to a control group [32]. The intervention aimed at a weight loss of >5% achieved by reduction of total dietary fat to <30% of calories from fat (<10% from saturated fat), >15 g per 1000 kcal fiber, and >30 min of exercise per day (walking, jogging, etc.). The subjects were instructed to eat whole grain products, vegetables, fruits, low-fat milk and meat products, and soft margarines frequently, and to avoid transfat and vegetables oils rich in mono-unsaturated fatty acids. Subjects attended seven counseling sessions with nutritionists during the first year of the intervention, and one session every 3 months in the subsequent years. The weight loss after 1 year was ~5 kg, and after 5 years the weight loss maintained was still 3 kg. This intervention produced a 58% reduction in the incidence of T2D. In a subsequent analysis weight loss was the strongest predictor of the diabetes preventive effect, and a weight reduction of 5% was associated with a reduction in relative risk of 61%. Every 3 kg of additional weight loss doubled this effect. In the

Figure 33.3 Weight loss outcomes for obese patients treated in weight management programs compared to predicted outcome of continued habitual lifestyle. It is noted that even weight stability may be a partial success and a goal in certain patients.
much larger, but very similar, American Diabetes Prevention Program, 3234 obese individuals (mean BMI 34 kg m⁻²) with IGT were randomized to intensive lifestyle, metformin, or a control group [31]. The goals of the lifestyle intervention were >7% weight loss, achieved by a diet providing <25% energy from fat and restricted to 1200–1800 kcal d⁻¹, and >150 min brisk walking per week [31]. The participants in the intervention group had 16 sessions with a nutritionist over the first 6 months, and subsequently one session per month for the remaining 2½ years. After 1 year the weight loss was 7%, and 5% was maintained throughout the trial as compared to the placebo group. The relative risk reduction of diabetes was the same as in the Finnish study—58% [31,32]. These studies strengthen the evidence that weight loss due to a combination of dietary intervention and physical activity is an effective strategy for diabetes prevention. Long-term follow-up of these studies show that a positive impact on diabetes risk and cardiovascular mortality is maintained even when the weight loss is not sustained.

**Weight loss management in type 2 diabetes**

Achieving and maintaining a desirable body weight is a major goal in management of T2D. Weight loss dramatically improves glycemic control, lipid profile, and blood pressure in obese individuals with T2D (ADA recommendations). For yet unknown reasons it is more difficult to produce and maintain weight loss in type 2 diabetics than in simple obese, but it is possible, and the health benefits are more substantial. In overweight and obese type 2 diabetics who have difficulty in losing weight on a conventional diet, a liquid-formula 800 kcal d⁻¹ diet may be a safe alternative. Irrespective of whether an intensive 800 kcal d⁻¹ diet consists of normal foods or of a liquid formula diet, weight loss after 3 months is 14–15% of initial body weight. This produces substantial improvements in glycemic control, blood lipids, and blood pressure. Most of the weight loss is maintained after 1 year. These studies show that it is possible to obtain substantial benefits in the treatment of type 2 diabetics through intensive diet therapy and lifestyle modification [27,33,34].

The large-scale Look AHEAD trial investigated whether intensive lifestyle intervention targeting weight loss in more than 5000 patients with T2D would reduce mortality over 13.5 years [33]. The study was discontinued because it was deemed unlikely to meet its primary objective, that is, to show survival benefits. Despite a rather modest weight loss the intervention very clearly demonstrated important quality of life benefits for participants, including improvement in HbA1c, but there was no reduction in hard cardiovascular end-points. However, the diabetic patients who lost weight benefitted from lowered HbA1c, reduced medications, increased physical fitness, less depression, as well as improved urinary incontinence, increased mobility, improved sleep apnea, and improved sexual function. Furthermore, participants who lost weight were more likely to experience remission of their T2D diabetes over the 4 years of the trial.

The lack of effect on CVD mortality may be attributed to the rather small weight loss of only 2.5% after 10 years, and the use of the traditional low-fat diet, of which any adverse effect on cardiovascular risk may have been masked by the weight loss.

**Physical activity in type 2 diabetes**

Physical activity alone may have a modest, but nevertheless important, effect on glycemic control. Favorable changes in glucose tolerance and insulin sensitivity usually deteriorate within 72 hours of the last exercise session. Consequently, regular physical activity is imperative to sustain glucose-lowering effects and improved insulin sensitivity.

According to a meta-analysis of intervention studies lasting >8 weeks addressing the effect of exercise on glycemic control and body weight in type 2 diabetics, aerobic and resistance training produced a beneficial effect on HbA1c of ~0.66% but no effect on body weight [35]. However, training may reduce body fat and increase lean tissue mass, so that BMI is unchanged. Physical activity has other important beneficial effects in type 2 diabetics, such as on cardiovascular risk factors and general well-being.

**Pharmacotherapy**

Current licensed pharmacotherapies vary in different regions of the world. In the USA they include phentermine hydrochloride (phentermine HCl), a sympathomimetic appetite-suppressant approved for short-term (up to 12 weeks) treatment of obesity in conjunction with dietary and lifestyle modifications, and orlistat, a gastric and pancreatic lipase inhibitor globally approved for the long-term management of obesity. In addition, lorcaserin, a 5HT-2c agonist, and a once-daily slow-release combination of phentermine and topiramate, are both approved by the US Food and Drug Administration (FDA) for chronic weight management as an adjunct to lifestyle modifications in obese adult patients, and in overweight (BMI >27 kg m⁻²) adult patients with at least one weight-related comorbidity. In Europe the European Medicines Agency (EMA) has only approved orlistat, and this compound is available in both a high-dose prescription version and in a low-dose over-the-counter version.

However, several other compounds, such as licensed antidiabetic medications with weight loss potential (acarbose, GLP−1 analogues, metformin) and compounds that are in phase III of development for the obesity indication, are relevant and the most important will be discussed briefly in this chapter.

**Phentermine**

Phentermine was approved by the FDA in 1959, and is the most commonly prescribed weight loss compound in the US [36]. It is only indicated for short-term treatment in patients with a BMI of >30 kg m⁻² or >27 kg m⁻² with obesity-related comorbidities (e.g. controlled hypertension, diabetes, dyslipidemia). Phentermine is a sympathomimetic related to amphetamine, and because of its abuse potential it is designated as a Schedule IV
controlled substance. Phentermine is administered orally, and the recommended daily dose is 15–37.5 mg in the morning, or one-half tablet two times daily. The dose should be individualized so that the lowest dose that produces a desirable weight loss is used.

**Mechanism of action**
Phentermine is an atypical amphetamine analogue that acts mainly by causing CNS release of norepinephrine (NE), thereby suppressing appetite and food intake. Phentermine shows only weak effects on dopamine and serotonin release, and is a weak uptake inhibitor of the serotonin transporter. Although increased NE release in the periphery may be perceived to increase the risk for adverse cardiovascular effects, NE released in CNS can act presynaptically to diminish sympathetic activity through a so-called clonidine-like effect [37]. This mechanism probably explains the finding that NE-uptake inhibitors (like sibutramine) paradoxically decrease sensitivity to sympathetic stimuli, such as stress and cold [37]. A slow-release phentermine preparation is also approved in the US for short-term treatment of obesity.

**Efficacy**
Phentermine has been evaluated in studies lasting up to 36 weeks, and a (30 mg d$^{-1}$) trial in overweight and obese women found weight loss of the order of 12–13 kg versus 4.8 kg in controls with both continuous and intermittent treatment regimens [38–40]. No long-term, randomized, placebo-controlled trial is available, but Hendricks et al. [41] found that patients treated with phentermine (15–37.5 mg d$^{-1}$) showed significantly greater weight loss over 2 years (12.7%) compared to a group on a low-carbohydrate ketogenic diet alone (8.4%). In a pooled analysis of studies lasting 2–24 weeks phentermine produced a 3.6 kg greater weight loss than placebo.

**Adverse effects**
The most common side effects are dry mouth and insomnia. Phentermine may be associated with cardiovascular complications, such as primary pulmonary hypertension, palpitations, tachycardia, increased blood pressure, and ischemic events, and is therefore contraindicated in patients with current or past cardiovascular disease, including uncontrolled hypertension. However, in patients in whom phentermine induces weight loss the treatment may actually lower blood pressure. Other contraindications are the same as those for monoamine oxidase inhibitors: hyperthyroidism, glaucoma, agitation states, pregnancy, or a known history of drug abuse. The abuse potential for phentermine is probably minor [42].

**Phentermine/topiramate**
Qsymia® (trade name in the USA) is a combination product for oral administration of immediate-release phentermine and extended-release topiramate licensed in the US as an adjunct to a reduced-calorie diet and increased physical activity for chronic weight management in obese patients (BMI ≥30 kg m$^{-2}$), or overweight patients (BMI ≥27 kg m$^{-2}$) with weight-related comorbidities such as hypertension, T2D, dyslipidemia, or central adiposity (abdominal obesity).

**Mechanism of action**
Phentermine mediates norepinephrine release in the CNS, and thereby induces weight loss by reducing food intake (see earlier). Topiramate exhibits a combination of properties that involves modulatory effects on sodium channels, enhancement of GABA-activated chloride channels, inhibition of excitatory neurotransmission through actions on kainite and AMPA receptors, and inhibition of carbonic anhydrase isoenzymes, but its precise mechanism of action to reduce food intake is not known. Topiramate was approved in 1996 for the treatment of seizures at doses up to 400 mg d$^{-1}$, in 2004 for the prevention of migraine headaches at doses up to 100 mg d$^{-1}$. Topiramate was initially also developed for the weight loss indication with 96 or 192 mg d$^{-1}$, and demonstrated long-term weight loss efficacy, but the program was stopped due to tolerability issues, that is, cognitive and speech disorders, and paresthesia [43]. The combination of phentermine and topiramate was developed for obesity management in three dosages taken once in the morning: low-dose (3.75 mg P/23 mg T), mid-dose (7.5 mg P/46 mg T), and high-dose (15 mg P/92 mg T), and the weight loss effects of its two components seems to be at least additive. It is recommended that doctors prescribing Qsymia follow a treatment algorithm: they should start with the low dose daily for 14 days; then increase to the mid dose, and then discontinue or escalate dose if 3% weight loss is not achieved after 12 weeks on this dose. They are also advised to discontinue Qsymia if 5% weight loss is not achieved after 12 weeks on maximum daily dose of 15 mg/92 mg. Discontinue 15 mg/92 mg dose gradually (as described) to prevent possible seizure. Do not exceed 7.5 mg/46 mg dose for patients with moderate or severe renal impairment or patients with moderate hepatic impairment.

**Efficacy**
The amount of weight loss achieved with combination therapy is of a greater magnitude than what could be achieved with either agent alone [43–47]. Data from phase 3 studies indicate that Qsymia produces significant weight loss in conjunction with improvements in obesity-related comorbidities such as hypertension, diabetes, and dyslipidemia. The three-dose regimens produce weight loss that is about 3, 7, and 9 kg greater than placebo plus diet and exercise program in 1-year trials (Figure 33.4). The corresponding weight loss in type 2 diabetics was ~7 kg [47]. In pooled analyses, Qsymia produced greater reductions in waist circumference, LDL-C, HbA1c, and SBP and DBP than placebo, and improvements in HDL-C compared to placebo. In pooled data of hypertensive subjects from the two 1-year studies, significant reductions in SBP (~9 mmHg) and DBP (~5 mmHg) were seen in all Qsymia groups, while heart rate was slightly increased in the 15/92 dose group (~2 bpm).
A large-scale cardiovascular hard endpoint study is ongoing. The primary endpoint will be the time from randomization to a nonfatal myocardial infarction, nonfatal stroke, or cardiovascular death.

**Adverse effects**
Several of the most commonly observed adverse events, occurring in 5% or more of study subjects, included paresthesia, dizziness, dysgeusia, insomnia, constipation, and dry mouth. Slight increase in heart rate, psychiatric and cognitive adverse effects, and metabolic acidosis are potential safety concerns. While the overall incidence rates are low, about 4–7 times as many subjects randomized to high-dose Qsymia versus placebo discontinued study participation due to anxiety-, sleep-, and depression-related adverse events. However, these effects are well-known and characterized side effects of one or the other component agent and do not represent novel side effects engendered through the combined pharmacology of the two drugs. There is no evidence of any new or unexpected safety issues with the combination relative to phentermine or topiramate monotherapy.

**Lipase inhibitors (orlistat)**

**Mechanism of action**
Malabsorption of dietary fat is an obvious drug target, and a number of compounds that act as intestinal lipase inhibitors have been investigated for use in the treatment of obesity, but only orlistat is currently approved and freely available. Orlistat is a specific inhibitor of intestinal lipase, the enzyme secreted from the exocrine pancreas and responsible for enzymatic fat digestion [48,49]. Orlistat inhibits lipase activity through formation of a covalent bond with serine within the catalytic site of gastric and pancreatic lipase. Orlistat is available on prescription at a recommended therapeutic dose (120 mg), and as an over-the-counter half-dose preparation (60 mg), both taken immediately prior to or within 1 hour of each of the 3 main daily meals. Fecal energy loss is not fully compensated by an increased caloric intake so orlistat produces a negative energy balance and weight loss. Orlistat is recommended for use as pharmacologic support for a nutritional strategy of a modest caloric restriction and a diet providing less than 30% of energy from fat. An obese patient with a typical energy requirement of 2600 kcal d⁻¹ will, on a 600 kcal deficit diet with 30% of calories from fat, experience an extra loss of 220 kcal fat in fecal energy.

**Efficacy**
The efficacy and safety of orlistat have been documented in several short- and long-term double-blind trials, some lasting for up to 4 years, involving a total of almost 30,000 patients, making orlistat the most well-investigated drug for the treatment of obesity. In most trials orlistat induces a dose-dependent
reduction in body weight, and the 120 mg dose induces a mean weight loss that typically exceeds that seen in the placebo group by 3–4 kg, to some extent independent of the degree of dietary energy restriction and other ancillary treatment. The weight loss typically continues for the first 3–6 months of treatment and then plateaus and patients may subsequently remain weight stable or regain some weight depending on the intensity of the dietary treatment. In obese subjects with the metabolic syndrome (insulin resistance syndrome) orlistat produces weight loss of the same magnitude as in simple obese, and the weight loss is associated with significant decreases in fasting insulin, triglycerides, the LDL/HDL cholesterol ratio and increase in HDL-cholesterol [50].

Use in prediabetes
Treatment of obese subjects with IGT by orlistat as an adjuvant to diet and lifestyle modification has been shown to decrease the rate of conversion to T2D. In a multicenter trial 3304 non-diabetic obese subjects underwent intensive lifestyle modification with a dietary energy deficit of 800 kcal d⁻¹, and were randomized to either orlistat or placebo for 4 years [51]. The mean weight loss was greater by 4 kg in the orlistat group after 1 year, and most of this weight loss was maintained throughout the rest of the study period. Weight loss was 6.9 kg in the orlistat group and 4.1 kg in the placebo group after 4 years. Nine percent in the placebo group but only 6.2% in the orlistat group had developed T2D after 4 years (relative RR = 37%). In addition, orlistat patients obtained greater reductions in waist circumference, LDL cholesterol, and systolic and diastolic blood pressures than placebo patients. The study clearly shows that the additional weight loss achieved by orlistat is sufficient to reduce the incidence of diabetes, even among more nonselected obese subjects where 80% had normal glucose tolerance. In the 20% of patients with IGT at enrollment the intensive lifestyle change alone resulted in a similar absolute diabetes incidence as the lifestyle treatment in the Diabetes Prevention Program [31].

Use in diabetics
One-year multicenter trials have been conducted with the use of orlistat in addition to each of the treatments of T2D: sulfonylurea, insulin, and metformin. Whatever medication the diabetics received at baseline, orlistat produced weight loss above diet and placebo of 2–5 kg [48], reductions in HbA1c of 0.46% (sulfonylurea), 0.35% (insulin) [52,53], and 0.29% (metformin). Greater reductions were also seen in LDL-cholesterol and blood pressure in the orlistat groups.

Adverse effects
Systemic absorption of orlistat and its metabolites is negligible, so adverse effects of orlistat are a consequence of the partial reduction of fat absorption, leading to an increased proportion of undigested fat in the bowel. The most frequent adverse effects of orlistat are flatulence, flatulence with discharge, oily spotting, fecal urgency and incontinence, oily stools, and steatorrhea. In most patients these adverse effects are self-limiting and mostly transient, generally occurring within the first months of treatment and often triggered by the ingestion of high-fat meals. However, the adverse effects tend to limit its use in many patients.

The unabsorbed fat binds some fat soluble vitamins (A, D, and E) and secondary nutrients (beta-carotene, lycopene, flavanoids, etc.). A simple vitamin supplement and increased intake of fruit and vegetables counteracts this effect. Pharmacodynamic studies suggest that orlistat does not affect the pharmacokinetic properties of lipid-soluble drugs such as digoxin, phenytoin, warfarin, glyburide, furosemide, captopril, nifedipine, atenolol, or oral contraceptives [54].

Lorcaserin
Mechanism of action
Lorcaserin is a selective serotonin 2C (5-HT₂C) receptor agonist that is intended for weight reduction and weight maintenance in obese patients, or overweight patients (BMI ≥ 27–30 kg m⁻²) who have one or more weight-related comorbid medical conditions [55–57]. Serotonin was found to decrease food intake and reduce body weight in animals 50 years ago, and these effects are mediated in part through activation of centrally located 5-HT₂C receptors. The historical weight control agents fenfluramine and dexfenfluramine, however, lacked specificity; each enhanced serotonin release and blocked its reuptake, leading to non-selective activation of multiple serotonin receptor subtypes [58]. As a result, fenfluramine caused serious adverse events that included heart valve vegetations that led to echocardiographically apparent aortic and mitral valve insufficiency. Serotonin valvulopathy appears to be mediated through activation of serotonin 2B (5-HT₂B) receptors located in the heart, whereas lorcaserin predominantly activates 5-HT₂C receptors without significant agonism of the 5-HT₂B receptor at therapeutic doses. The agonist activity of lorcaserin at the 5-HT₂A receptor, which has been linked to mood and perceptual effects, is minimal. Given that 5-HT₂C receptor expression is primarily limited to a few regions of the central nervous system, lorcaserin should cause weight loss with few unintended pharmacologic effects.

Efficacy
Lorcaserin caused significant, dose-dependent weight loss in obese patients and overweight patients with at least one weight-related comorbid condition. Nondiabetic patients treated with lorcaserin 10 mg BID lost 5.8 kg over 1 year as compared to 2.5 kg in the placebo group [59]. A meta-analysis found weight loss of 3.2 kg (95% CI 2.7–3.8) greater than with placebo in trials of 1-year duration [60]. The use of lorcaserin for 8 and 12 weeks reduced weight by 1.6 kg and 2.9 kg, respectively. Of patients treated with lorcaserin 10 mg 47.1% achieved ≥ 5% reduction in body weight after 52 weeks as compared to 22.6% of patients treated with placebo, and 22.4% of patients treated with lorcaserin achieved ≥10% weight loss as compared to 8.7% in the placebo group.
In comparison to placebo, lorcaserin decreased waist circumference, blood pressure, total cholesterol, LDL-cholesterol and triglycerides; however, it did not statistically affect heart rate or HDL-cholesterol [56], and also produced positive effects on glycemic control and insulin resistance, inflammatory markers, and quality of life in trials of 1-year duration, and many of these effects were maintained in year 2.

In patients with T2D lorcaserin produced a weight loss of 4.7 kg as compared to 1.6 kg in the placebo group, and it improved glycemic control relative to placebo and reduced the use of antihyperglycemic medications. Over half of patients taking lorcaserin achieved the ADA recommended glycemic goal of HbA1c ≤7% compared to 26% in the placebo group.

Adverse effects

The most frequent adverse effect is headache, but also dizziness, nausea, fatigue, urinary tract infections, dry mouth, and back pain are frequent side effects of lorcaserin. The majority of headache, dizziness and nausea effects occur early on in the course of treatment, tend to be mild to moderate, and generally do not recur after resolution of the initial event. In type 2 diabetes hypoglycemia is the most frequently reported adverse event. Though lorcaserin has been associated with excess neoplasia or with evidence of hepatic toxicity in some rodent studies, there has been no evidence of this in humans. Lorcaserin does not cause clinically significant increases in serum prolactin, nor does it appear to increase the risk of cardiac valvular disease. Lorcaserin should not be used in patients with creatinine clearance ≤30 mL min⁻¹.

Pharmaceutical compounds for diabetes treatment with weight loss properties

Metformin

Metformin is a biguanide that is approved for the treatment of T2D. It lowers plasma glucose by reducing intestinal absorption, suppressing hepatic glucose production, and increasing peripheral insulin sensitivity. It has long been known to induce a slight weight loss not only in diabetics but also in obese subjects with metabolic syndrome or IGT. The size of the placebo-subtracted weight loss induced by metformin in diabetics varies from 0.5 kg to 8 kg, probably depending on BMI, age, and glycemic control [53]. In the Diabetes Prevention Program of obese subjects with IGT, metformin produced a 2.0 kg weight loss and reduced the incidence of diabetes by 31% above placebo after 3 years [31]. A subgroup analysis revealed no difference in efficacy across gender and ethnicity, but the effect was mainly seen in subjects aged 25–59 years with a BMI above 30 kg m⁻², whereas it was largely ineffective in nonobese (BMI < 30 kg m⁻²) and those older than 60 [31]. In this trial more than 1000 patients entered each treatment arm and the resulting high statistical power gives more credit to the outcome than those of smaller trials.

Metformin may produce weight loss through effects on AMPK and mitochondrial function, and the improvement in insulin sensitivity may increase the central effect of insulin where it acts as a satiety hormone. Another possibility is that reduced food intake is caused by nausea and other gastrointestinal effects. There are no indications of any thermogenic effect of metformin.

Amylase inhibitors (acarbose)

Inhibition of digestion of starch and disaccharides reduces absorption of di- and monosaccharides and produces a reduction in blood glucose and a negative energy balance, providing that energy intake is not increased by counterregulatory mechanisms. Acarbose is an antidiabetic drug that lowers blood glucose by inhibition of α-glucosidase in the gastrointestinal tract, thereby delaying the hydrolysis of ingested disaccharides and complex carbohydrates. Theoretically, efficacy with respect to body weight regulation is modest, as starch that is not digested in the small intestine is fermented in the colon and finally absorbed as short chain fatty acids. It is estimated that fermentation of carbohydrates in the colon is associated with a 50% reduction of energy content available for the body, thus reducing the energetic value of carbohydrates from 4 to 2 kcal g⁻¹.

However, the effect of acarbose on glycemic control in diabetics is modest, and the effect on body weight is negligible in simple obese [61], obese with IGT [62], and in type 2 diabetics [63]. In a weight maintenance study in simple obese subjects Hauner et al. failed to find any effect on body weight of 6 months acarbose treatment following an initial diet-induced weight loss of 10 kg [61]. Three years of acarbose treatment produced a weight loss of 0.8 kg compared to placebo, and the risk of progression to diabetes over 3.3 years was reduced by 25%, which could be attributed mainly to the weight loss [62]. However, 31% of the patients in the acarbose group, compared to 19% in placebo group, discontinued treatment early, mainly due to gastrointestinal side effects [62]. About 50% more patients reported side effects on acarbose than on placebo, mainly flatulence, abdominal pain, and diarrhea. The effects of acarbose in T2D are equally unpromising. In the UKPDS, 1946 type 2 diabetics randomized to acarbose or placebo were followed for 3 years [63]. Fifty-eight percent of the patients discontinued acarbose versus 39% drop-out in the placebo group. In an intention-to-treat analysis acarbose produced only a 0.2% lower HbA1c than placebo, while in those who remained on the allocated therapy the HbA1c was 0.5% lower. Mean body weight was lower by 0.4 kg in the acarbose group after 1 year, but no difference was seen after 2 or 3 years [62].

In conclusion, acarbose produces only a very modest (<1 kg), if any, weight loss and, taken together with its low tolerability due to prevalent gastrointestinal side effects, has no place in the treatment of obesity. Acarbose may, however, have some role in type 2 diabetics with a capacity to tolerate the compound [63].
GLP−1 analogues

The GLP−1 analogues exenatide (two injections per day or long-acting release form once weekly) and liraglutide (one injection per day) are licensed for the treatment of T2D, and more are under investigation (albiglutide, taspoglutide, lixisenatide [64] and more). Some of these analogues are being developed for both diabetes and obesity indications [65].

Mechanism of action

The first evidence that GLP−1 is a hormone responsible for inducing postprandial satiety and reduced spontaneous caloric intake in normal humans was published in 1998 [66]. It was also found that obese individuals require higher doses than normal weight individuals to achieve the same degree of satiety, and that the satiety effect requires higher dosages than those required to obtain glucose lowering. Native GLP−1 has a very short half-life but development of analogues has enabled treatment with injections producing supraphysiologic plasma levels. Exenatide, for TID, and liraglutide once daily, are both licensed for the management of T2D as they produce a clinically relevant reduction in HgbA1c. Because the satiety action of GLP−1 is further stimulated by higher doses than the glucose-lowering effect, current doses of exenatide and liraglutide (i.e., 1.2 and 1.8 mg d−1) used in T2D are not optimal for weight loss.

Efficacy

The weight loss in type 2 diabetics in RCTs of GLP−1 agonists is typically only 1–3 kg greater than placebo over 6 to 12 months, with no difference between short- or long-acting agents, or between exenatide and liraglutide. Liraglutide has been assessed in a number of long-term trials in nondiabetic obese populations at doses up to 3.0 mg d−1. Astrup et al. reported a dose-finding study with different doses of liraglutide versus placebo and versus orlistat in obese subjects treated with a mild hypocaloric diet over 24 weeks [67]. Liraglutide produced a dose-related weight reduction that was greater than placebo for all doses from 1.2 to 3.0 mg d−1, with a clear separation of dosages. Liraglutide 3.0 mg produced mean weight loss 4.4 kg greater than placebo, and 3.1 kg greater than orlistat. This trial was extended to 1 year, and subsequently all liraglutide and placebo patients switched to liraglutide 2.4 mg, then to 3.0 mg [68]. From randomization to year 1 liraglutide 3.0 mg recipients lost 5.8 kg (3.7–8.0) more weight than those on placebo, and 3.8 kg (1.6–6.0) more than those on orlistat (Figure 33.5). In the SCALE program a number of trials have extended the efficacy of 3 mg liraglutide in nondiabetic obese patients, and placebo-adjusted long-term weight loss has been in the order of 5–7 kg.

Liraglutide exerts beneficial effects on almost all clinically relevant cardiovascular risk factors, that is, blood pressure, blood

![Figure 33.5](image-url) Two-year weight loss effects of the GLP−1 agonist liraglutide. Change in body weight from screening over 2 years, presented as observed data for individuals completing each scheduled visit. A randomized, double-blind, placebo-controlled 20-week study with 2-year extension in 564 obese patients. Participants received diet (500 kcal deficit per day) and exercise counseling during 2-week run-in, before being randomly assigned to once-daily subcutaneous liraglutide (1.2, 1.8, 2.4, or 3.0 mg), placebo, or open-label orlistat (120 mg x 3). After 1 year, liraglutide/placebo recipients switched to liraglutide 2.4 mg, then 3.0 mg. At year 2, participants on liraglutide 2.4/3.0 mg for the full 2 years lost 3.0 kg more weight than those on orlistat. Completers on liraglutide 2.4/3.0 mg maintained a 2-year weight loss of 7.8 kg from screening. Source: Astrup et al. 2012 [68]. Reproduced with permission of Nature Publishing Group.
lips, inflammatory markers, and glucose metabolism in non-diabetic obese patients. In the trial by Astrup et al. 31% of the obese patients had prediabetes at randomization, and 41% met the criteria for metabolic syndrome. Between 52–62% of the liraglutide-treated individuals with prediabetes at randomization achieved normal glucose tolerance at year 2, compared with 26% of those on orlistat.

Safety of GLP−1 analogues
The most frequent side effects are mild to moderate transient nausea and vomiting. These side effects are clearly dose-related, and it is likely that individual differences in sensitivity to GLP−1 are responsible, so that increasing receptor stimulation may be responsible for moving from enhanced satiety to nausea. This interpretation is supported by findings showing that weight loss with liraglutide 3.0 mg at 1 year from randomization was 10.0 kg for those with nausea and/or vomiting, and 7.1 kg for those without (difference 2.9 kg (0.5–5.3, p < 0.02)). In the Astrup et al. trial 53% of patients taking liraglutide 3.0 mg reported nausea compared to 8% in the placebo group and 7% in the orlistat group [68]. Most nausea/vomiting episodes started in the first 6 weeks of treatment, were transient, and were of mild or moderate intensity. Few episodes were serious. Another cardiovascular side effect is an increase of 1–4 bpm after 1 year intervention. This is presumably caused by a direct stimulation of the sinus node of the heart.

Patients with T2D have a 2.8-fold, and obese subjects a 2.2-fold, higher risk of developing pancreatitis than the general population [69,70]. A possible mechanism for some of the cases of pancreatitis observed in the trials may be the triggering of acute pancreatitis among susceptible individuals, such as obese and type 2 diabetic subjects, by weight-loss induced gallstones. A recent meta-analysis concluded that there is no evidence to support an increased risk of acute pancreatitis associated with the use of GLP−1 agonists [71] but we are still waiting for the outcome of some of the long-term outcome studies like LEADER, ELIXA, and others before this potential side effect can be quantified.

Other pharmacologic agents under development

Bupropion/naltrexone
Mechanism of action
Bupropion and naltrexone were approved for other indications in 1984−85, but both have been associated with weight loss. Bupropion stimulates hypothalamic pro-opiomelanocortin (POMC) neurons, with downstream effects to reduce food intake and increase energy expenditure. Naltrexone blocks opioid receptor-mediated POMC autoinhibition, augmenting POMC firing in a synergistic manner. Given the fact that both drugs have effects on addiction, alcohol and nicotine, respectively, the combination was considered to induce weight loss through sustained modulation of CNS rewarding pathways [72].

Efficacy
A number of trials with the sustained release combination of naltrexone and bupropion, in combination with mild diet and exercise in overweight and obese subjects, have shown good efficacy in terms of weight loss as the combination produced a weight loss around 4−5% greater than placebo [73,74]. Bupropion is associated with a slight increase in blood pressure and heart rate, whilst the phase III trials on the combination have found unchanged heart rate and blood pressure despite the reduction in body weight.

Adverse effects
Bupropion is generally considered a reasonably safe medication for smoking cessation. However, 30−40% of treated patients may suffer from insomnia, and generalized seizures occur in approximately 0.4%. Other side effects are agitation and anxiety [74]. Naltrexone has been associated with adverse events such as difficulty in sleeping, anxiety, nervousness, and headache. The most frequently reported treatment-emergent adverse events for the combination compound are nausea, insomnia, headache, and constipation, but no tendencies to worsening of various anxiety, depression, or suicidal subscale scores have been found, nor were there more adverse events or withdrawals due to these psychiatric conditions. These findings suggest that the combination may have fewer adverse psychiatric effects than its components given separately.

References


Chapter 33


CHAPTER 34

Treatment of obesity: bariatric surgery

Karl J. Neff1, Ele Ferrannini2 and Carel W. le Roux1

1Diabetes Complications Research Centre, Conway Institute, University College Dublin, Dublin, Ireland
2Department of Clinical and Experimental Medicine, University of Pisa, Pisa, Italy

Key points

• Bariatric surgery is the most effective treatment for obesity.
• In addition to weight loss, bariatric surgery has significant effects on metabolic disease, and particularly diabetes.
• These surgical procedures should now be considered as treatments for metabolic disease, to be added to medical therapy in selected cohorts.
• There are multiple potential mechanisms underlying weight loss and improved glucose homeostasis after each procedure, which likely act in concert.

Introduction

The primary focus of bariatric surgery since its introduction has been the reduction of body weight [1–3]. However, as well as sustained effects on weight loss, bariatric surgery is associated with improvements in several obesity-related conditions, from airway diseases to cancer [1,2,4]. Despite increasing evidence of benefit for obesity-associated diseases, the major national guidelines in the United States and Europe for selecting candidates for bariatric surgery continue to focus on body mass index (BMI) as a primary indication, and outcomes are based on changes in weight. This standpoint is changing, and the potential benefits of bariatric surgery beyond that of weight loss are beginning to be appreciated.

In metabolic disease, bariatric surgery results in marked improvements in glycemic control in type 2 diabetes mellitus (T2DM). In some patients, remission of diabetes, hypertension, and dyslipidemia can be achieved within weeks, and in the case of diabetes this can be before the onset of significant weight loss [5]. However, the definition of remission of these metabolic disorders has varied in the past, in the absence of a consensus on how this term should be defined. In most studies, remission is defined by clinical report rather than objective examination or biochemical analysis. The discontinuation of pharmacologic therapy for metabolic disorders has also been used as a surrogate for remission. This means that these studies should be interpreted with some degree of caution.

Findings from randomized controlled trials have demonstrated that bariatric surgery is superior to best medical care alone for obesity-associated T2DM [6,7]. Bariatric surgery improves glycemic control for up to two years in randomized controlled data, but also reduces mortality and maintains adequate control of diabetes for at least 10 years in observational data [1,2,6,7]. The remit of surgery is now beyond that of weight control or reduction, and surgery could be considered a useful therapy in more patients with metabolic disease or dysfunction.

Current surgical approaches

The most commonly performed bariatric procedures are Roux-en-Y gastric bypass (RYGB), adjustable gastric banding (AGB), and vertical sleeve gastrectomy (VSG) [8]. Biliopancreatic diversion with or without duodenal switch (BPD, BPD-DS) is less commonly performed [9]. All procedures can be performed laparoscopically with a low rate of complications such as wound infection and incisional hernias [10].

RYGB involves the division of the stomach into an upper gastric pouch, which is ~30 mL in volume, and a lower gastric remnant (Figure 34.1(a)). The gastric pouch is anastomosed to the jejunum after it has been divided some 30–75 cm distal to the ligament of Treitz; this distal part is brought up as a “Roux limb.” The excluded biliary limb including the gastric remnant is connected to the bowel some 75–150 cm distal to the gastrojejunostomy (Figure 34.1(b)).

AGB, vertical banded gastroplasty (VBG), and VSG have a surgical focus on the stomach, with either a partial gastrectomy (VSG) or gastric restriction by means of bands (AGB) or stapling of a portion of the stomach (VBG). Despite this focus,
in AGB a silastic band is applied around the stomach just below the gastroesophageal junction, and is tightened through a subcutaneous access port by the injection or withdrawal of fluid (Figure 34.2). In VSG the stomach is transected vertically thereby creating a gastric tube and leaving 100–200 mL-size pouch (Figure 34.3). VSG has been shown to have metabolic effects comparable to RYGB [11]. VBG is a procedure that is not commonly performed currently, but which was used in landmark studies such as the Swedish Obese Subjects (SOS) study. In this procedure, the superior stomach is stapled vertically to create a pouch which is then separated from the rest of the stomach with a band.

BPD comprises a partial gastrectomy with formation of a 400 mL gastric pouch [12]. The small bowel is divided 250 cm proximal to the ileocecal valve, and the alimentary limb is connected to the gastric pouch to create a Roux-en-Y gastroenterostomy. An anastomosis is performed between the excluded biliopancreatic limb and the alimentary limb 50 cm proximal to the ileocecal valve (Figure 34.4) [13].

There are several endoscopic techniques in development that aim to replicate some of the metabolic effects of surgery [13]. Endoscopically placed synthetic duodeno-jejunal bypass liners such as the EndoBarrier® have been developed as an alternative to bariatric surgery and can result in a mean weight loss of 10–15% [14]. Complications such as sleeve migration and bowel obstruction can occur with the current technology, and the device produces gastrointestinal side-effects in many recipients. Given these limitations and the lack of long-term data, the role for devices such as EndoBarrier® remains to be determined.

**Weight loss after bariatric surgery**

The dominant effect of bariatric surgery, which is probably responsible for the majority of the metabolic benefits at least
in the longer term, is sustained weight loss [9,10,15,16]. Depending on the type of surgery, maximal weight loss usually occurs 12 months postoperatively, with peak total weight loss of 33% after RYGB [17]. While RYGB results in greater weight loss than AGB in most studies, high-quality postoperative care after AGB can result in weight loss comparable to RYGB [18,19]. There are no differences in weight loss after RYGB and VSG 36 months postoperatively [20]. Longer term data on VSG are scant at present, but the data that are available suggest that significant weight regain may be frequent [21]. While BPD results in greater weight loss than RYGB, the complication rate is also higher [22,23]. Despite greater weight loss, metabolic and quality of life outcomes are comparable between RYGB and BPD in patients with BMIs above 55 kg m$^{-2}$ [22,23].

In a minority, peak weight loss can be met by 6 months postoperatively, or may not be achieved until 24 months postoperatively (16). However, some weight regain is common thereafter and the mean 10-year weight reduction is 25% for RYGB and 13% for AGB [1]. Weight regain is described after all major procedures, but the etiology of regain is complex and multifactorial. Factors such as follow-up protocols can be as important as the type of procedure, and many patient-specific factors can influence weight outcomes significantly after RYGB and AGB [24]. Longer term data on VSG are still awaited and this represents a reason for caution with this relatively new procedure [21].

**Mechanisms of weight loss**

RYGB was initially designed to combine malabsorption and restriction. However, using markers such as serum albumin and fecal fat, RYGB does not appear to result in malabsorption [25,26]. After RYGB, the reduction in combustible energy absorption is moderate (6–11%) [25]. Therefore, malabsorption is not usually a major mechanism of weight loss. Malabsorption may become more important in those with shorter lengths of the intestinal limbs, or in those who maintain high-fat diets after RYGB.

BPD has a greater malabsorptive effect than RYGB, and BPD recipients can consume over 3000 kcal daily and still maintain long-term weight loss [27]. Increased fecal fat, steatorrhea, and hypoalbuminemia after BPD are common [27]. This is consistent
Figure 34.4 Biliopancreatic diversion: the combination of sleeve gastrectomy and intestinal rearrangement results in significant changes to the metabolism, but can result in malabsorption. (For a color version of this figure, please see the color plate section.)

with malabsorption, suggesting that this is a significant contributor to weight loss after BPD. Fecal caloric density is not altered by VSG in animals but there are no data in humans [28]. VSG and AGB are not considered to produce significant macronutrient malabsorption.

AGB reduces stomach volume, and food within the smaller stomach pouch results in early gastric distension [29]. However, there is a lack of compensatory, high calorie seeking behavior in the majority of patients after AGB, which might be expected in the context of reduced food volume. The rapid weight gain seen after reversal of AGB suggests a physiologic attenuation of appetite that is modulated by the band [30]. Indeed, changes in appetite are reported following AGB [31–33]. Gut hormones that induce satiety, such as glucagon-like-peptide 1 (GLP-1) and peptide YY (PYY), are not altered by AGB, and inhibition of gut hormone responses with octreotide (a somatostatin analogue) does not affect food intake in AGB recipients [32,33]. Therefore, it is likely that AGB exerts its effects on satiety by neural signalling, probably through the vagal nerve and its branches [30]. This is supported by data that show increased satiety associated with increased pouch pressure [34]. Pouch emptying rates and changes in pouch pressure are not associated with satiety, but the dilatation or pressure effect on the gastric pouch produces a satiation effect in the absence of any consistent change in gut hormones [33].

The mechanisms to explain weight loss after VSG are not fully elucidated. Gastric emptying and intestinal transit appear to be faster after VSG unless the antrum is preserved [35–39]. The generation of high intraluminal gastric remnant pressures and the disruption of the gastric pacemaker at surgery may explain these effects, and why anorexigenic gut hormones such as GLP-1 and PYY are elevated after VSG in magnitudes similar to RYGB. More studies examining the mechanisms behind VSG are required.

RYGB reduces stomach volume, which may produce early gastric distension and early satiety resulting in reduced meal size [29]. We would expect that reduced meal size volume results in a compensatory increase in appetite for calorie-dense food. However, RYGB patients report reduced hunger, increased satiety, and lower consumption of energy-dense foods postoperatively [30]. Therefore, it is not simply a matter of gastric pouch size. Randomized controlled trials have shown that the similarly “gastro-restrictive” procedure of VBG, results in less weight loss and less change in food preferences as compared to RYGB [40]. Hormonal and neural signals likely play important roles in this effect.

Satiety: the role of gut hormones and neural signals

Increased satiety and decreased hunger occur within days following RYGB, and this is associated with changes in the postprandial levels of gastrointestinal hormones that induce satiety, such as GLP-1 and PYY [31–33]. Postprandial levels of PYY and GLP-1 are increased from the second postoperative day after RYGB, prior to significant weight loss. Stimulated PYY and GLP-1 levels correlate with the amount of weight loss [31–33]. Inhibition of PYY and GLP-1 responses with a somatostatin analogue (octreotide) in patients after RYGB increases food intake, suggesting an important role of gut hormones in mediating enhanced satiety after RYGB [32].

Vagal nerve afferents are activated by the presence of nutrients in the upper gastrointestinal tract, and the perioperative preservation of vagal fibers results in greater sustained body weight loss after RYGB [41]. Thus, the rapid entry of food from the esophagus, through the small gastric pouch and the larger gastrojejunostomy, may trigger neural signals in the alimentary limb, which may contribute to long-term weight maintenance after RYGB [42,43].

These neural signals interact with gut hormone signals centrally to affect eating behavior, in association with other cues from the available food resources. These can be termed internal signals (such as gut hormones and vagal-mediated neural signals) and external signals (food cues, palatability, availability). The orbitofrontal cortex, hypothalamus, brainstem, and corticolimbic areas process sensory information and then
coordinate energy homeostasis by modulating food searching, sensing, and reward. Higher cortical centers implicated in emotional responses can influence food intake to provide reward, and this intake can exceed the requirements of energy homeostasis [44]. Dysfunction of this reward network may contribute towards dysfunctional eating and the development or maintenance of obesity [45].

Postoperatively, RYGB recipients tend to reduce meal size but increase meal frequency [46]. In randomized controlled studies, VBG results in a higher proportion of fat and carbohydrates in the diet compared to RYGB recipients, who prefer fruit and vegetables and have a reduced preference for high-fat foods [40]. Compared to VBG, RYGB patients tend to avoid calorie-dense foods, and prefer high glycemic index foods [47]. In addition, the reward areas of the brain are less active when presented with high caloric food after RYGB [48,49]. Therefore, RYGB remediates some “reward dysfunction,” and hormonal or neural signals may be implicated. This control mechanism remains to be fully elucidated, but is likely to be important in the effects of RYGB on weight and metabolic health.

**Energy expenditure**

In rodents, energy expenditure increases after RYGB despite reduced calorie intake [50,51]. In humans, the data are scanty and inconsistent, partly as a result of the inherent difficulty in measuring energy balance in humans [52–58]. Chronic calorie restriction normally produces a decrease in resting energy expenditure [50] and diet-induced thermogenesis is impaired in insulin-resistant subjects [59]. Therefore, increased insulin sensitivity is expected to enhance diet-induced thermogenesis [54]. Energy expenditure reflects the mass of metabolically active tissue, resting energy expenditure generally must be reduced postoperatively due to the loss of lean body mass along with the loss of fat mass. Whether such reduction is strictly proportional to the loss of lean tissue in all patients and with any surgical procedure is uncertain. The following diet-induced weight loss energy expenditure is less than predicted by the loss of lean mass and can contribute to weight regain [60]. Finally, in weight-reduced subjects who increase their physical activity, the associated increase in energy dissipation may be important, although moving a substantially leaner body costs less energy [55]. The postoperative changes in energy expenditure are likely to be the algebraic sum of all of these changes, and can therefore produce great variability within and between subjects.

**Treatment of diabetes and related metabolic diseases**

The current patient selection criteria for bariatric surgery focus on BMI, with a value of 40 kg m$^{-2}$ as the threshold for surgery in those without recognized comorbidities. Comorbidities, including metabolic disorders such as diabetes, lower the BMI threshold for surgery to 35 kg m$^{-2}$ [61,62]. Bariatric surgery for individuals with a BMI less than 35 kg m$^{-2}$ with obesity-related comorbidities is currently an area of intense investigation, but is not generally recommended [63].

Data in patients with BMI less than 35 kg m$^{-2}$ show that bariatric surgery can produce weight loss that brings them into the nonobese range, with associated improvements in fasting glucose levels and lipid profiles [64]. These data are mainly based on studies of RYGB and BPD [64]. In subjects with diabetes, over 80% achieve a HbA1c of less than 7% (53 mmol mol$^{-1}$) with a 3% major complication rate and no mortality [64]. This effect can be durable, with prospective data reporting 88% that RYGB recipients with a BMI of 30–35 kg m$^{-2}$ maintain a HbA1c less than 6.5% (48 mmol mol$^{-1}$) 6 years postoperatively [65].

RYGB improves glucose homeostasis even in those with BMIs less than 35 kg m$^{-2}$, a mean duration of T2DM of 11 years, and suboptimal glycemic control with a mean HbA1c of over 9% (75 mmol mol$^{-1}$) [66]. Over a 1-year follow-up period, over 80% of this cohort achieved HbA1c levels less than 7% (53 mmol mol$^{-1}$) and remained off antidiabetic therapy without excessive weight loss [66]. These data show that bariatric surgery could be considered in those with a BMI less than 35 kg m$^{-2}$ in the presence of diabetes, especially in those who are not meeting their treatment targets despite intensive medical therapy. Based on a recent systematic review, bariatric surgery in such candidates would improve HbA1c from high-risk ranges to near-normal ranges [67].

This has led the IDF to recommend that bariatric surgery be considered for treating T2DM in patients who are not adequately responding to medical treatment with a BMI of 30 kg m$^{-2}$ or greater [68]. This is a development of the metabolic surgery paradigm. In addition to the well-documented effects on glucose homeostasis, there are emerging data to suggest that bariatric surgery may have a role in the treatment of diabetic complications such as diabetic kidney disease [69,70].

**Diabetes remission**

Diabetes remission is widely discussed as an outcome of bariatric surgery, but remission following bariatric surgery is often transient [1,2]. Recurrence of T2DM is associated with longer durations of T2DM preoperatively, less weight loss postoperatively, and postoperative weight regain [71]. Recently, it has been recognized that remission may be an unhelpful goal, and that surgery should instead be recognized as another method for achieving long-term glycemic control in patients in whom less invasive therapies have failed. A major difficulty in interpreting these data are the differing definitions of remission, and differing time frames in study follow-ups. In the Swedish Obese Subjects study, approximately 50% of those who entered remission at 2 years after surgery had recurrence of diabetes at 10 years, although they were still left with much better glycemic control than presurgery [1]. After 6 years, RYGB reduced HbA1c to less than 6.5% (48 mmol mol$^{-1}$) in 62% of patients when used in conjunction with best medical therapies [72].
This differs from the current American Diabetes Association definition of diabetes remission as a HbA1c less than 6% (42 mmol mol\(^{-1}\)) and a fasting glucose less than 5.6 mmol L\(^{-1}\) off glycemic therapy [73]. This definition is controversial as many believe that it encourages clinicians to discontinue medication instead of using surgery as an add-on therapy to current medical therapy [73].

In the studies above, remission of diabetes was not a primary outcome when the participants were selected for bariatric surgery [16–18]. The candidates had a wide range of “metabolic phenotypes” [16–18]. Certain metabolic characteristics determine the likelihood of remission after surgery, and duration of diabetes and level of glycemic control preoperatively are important factors. Those with a greater duration of diabetes, higher preoperative HbA1c values, insulin dependence, and less weight loss are less likely to achieve postoperative remission of diabetes after RYGB or VSG [74]. Heterogeneous clinical characteristics of the cohorts with diabetes in the available studies, and ongoing changes in the definition of diabetes remission during the period of these studies, make interpretation of the data more difficult. Nevertheless, using the current American Diabetic Association criteria (a HbA1c less than 6% (42 mmol mol\(^{-1}\)) and a fasting glucose less than 5.6 mmol L\(^{-1}\) off glycemic therapy, diabetes remission occurs in at least 34% of all recipients of RYGB, VSG, and AGB at a mean follow-up of just under 2 years [73]. RYGB has greater efficacy in this regard, with a remission rate of over 40%, as compared to 26% of VSG and 7% of AGB, but selection bias may heavily influence these data, as may greater weight loss when compared to AGB [73]. Future studies on glycemic control should apply the same definitions to allow comparison between results.

The evidence to date supports that surgery should be considered as an add-on therapy for diabetes, where it can be expected to induce transient remission in those with shorter duration of diabetes and better glycemic control preoperatively. Remission as a goal of treatment would preselect candidates on the basis of the duration of their diabetes, and on the range of their hyperglycemia, rather than focusing on patients where less invasive therapies have failed and where surgery could add the most benefit when used with best medical care. The patients with a longer history of diabetes and uncontrolled hyperglycemia are less likely to remit, but may also be more likely to benefit as they are at a much higher risk of developing micro- and macrovascular complications. This would suggest that diabetes remission should not be the objective. Instead, it could be argued that the clinical objective should be to improve glycemic control in those with advanced metabolic disturbance, as evidenced by uncontrolled hyperglycemia, or those either at high risk of diabetic complications or with diagnosed diabetic vascular disease.

**Improving glycemic control with bariatric surgery**
Bariatric surgery with medical therapy produces greater glycemic control compared to best medical therapy alone, or surgery without best medical therapy [6,7,75]. In randomized controlled trials, RYGB, BPD, VSG, and AGB are superior for improving glycemic control than medical therapy alone [6,7,76]. In adults with BMIs of 30–40 kg m\(^{-2}\) and a duration of T2DM of less than 2 years, AGB can result in over 70% or patients with T2DM achieving a HbA1c of 6.2% (44 mmol mol\(^{-1}\)) or less, as compared to less than 15% of those on medical therapy attending a diabetes multidisciplinary team for review every 6 weeks [76]. Note that the improved glycemic control was in the context of a 10-fold greater weight loss in the AGB group as compared to the medical group over a 2-year follow-up period [76].

In another randomized controlled trial comparing BPD, RYGB, and medical therapy in patients with a BMI of 35 kg m\(^{-2}\) or more and T2DM for at least 5 years, 95% of BPD recipients achieved HbA1c ≤ 6.5% (48 mmol mol\(^{-1}\)) compared to 75% of RYGB recipients and 0% of the medical treatment group over 2 years [7]. Participants in the intensive medical arm attended a diabetes multidisciplinary team every 12 weeks for optimization of their diet, lifestyle, and medical therapy [7]. Over the 2-year study period the intensive medical therapy group lost nearly 5% of their baseline weight, compared to over 30% weight loss in both surgical groups [7]. In other randomized controlled data testing RYGB and VSG against intensive medical therapy administered along American Diabetes Association guidelines, a HbA1c of 6% (42 mmol mol\(^{-1}\)) was achieved by 42% of the RYGB group combined with best medical care, 37% of the VSG group combined with best medical care, and 12% of the medical group [6]. The cohort were obese, had T2DM for a mean duration of over 8 years, and were not meeting their current glycemic targets [6].

These studies were the first to use bariatric surgery as “add-on therapy” to best medical care. Metformin, lipid pharmacotherapy, and renin-aldosterone-angiotensin system antagonists were continued after surgery where possible. The study cohorts represent those with significant metabolic disturbance refractory to medical therapy, who are the patient cohort that struggles to achieve safe treatment targets with nonsurgical treatments. These data demonstrate that in a selected group, surgery combined with best medical care can be the optimal therapeutic approach. However, long-term data is needed before making final judgment. Glucose homeostasis worsens in the first 2 postoperative years after bariatric surgery [1,68]. Therefore, longer follow-up of these cohorts is necessary to ensure that the effects are durable. Further studies on cost–benefit and risk–benefit ratios should be performed to ensure that the complications and cost inherent in bariatric surgery do not outweigh the benefits in this group.

**Bariatric surgery and microvascular complications**
The improved glycemic control associated with bariatric surgery could be expected to reduce the burden of vascular complications in diabetic cohorts. So far there have been no randomized controlled trials examining the effect of bariatric
surgery on microvascular outcomes; however, observational data support the hypothesis that bariatric surgery may have a role in managing not only diabetes, but also its complications.

Most of the available data on microvascular diabetic complications are on diabetic kidney disease. In a retrospective analysis of 25 RYGB patients, renal function improved from an impaired baseline estimated glomerular filtration rate (eGFR) of 48 mL min⁻¹ 1.73 m⁻² to over 60 mL min⁻¹ 1.73 m⁻² [77]. The difficulty with these type of data is that they may be falsely reassuring as any weight loss will reduce serum creatinine; thus, the improved eGFR estimation based on the Modified Diet in Renal Disease (MDRD) formula may reflect the reduction in lean body mass rather than improvements in renal function. Some studies have used other measurement techniques of GFR such as creatinine clearance and urinary cystatin C:creatinine ratio, and found that RYGB improves renal function in cohorts with or without diabetes [78]. There are few studies in which the effect of the surgical procedure itself on renal function can be assessed independently of weight loss. This is an area that needs further specific investigation.

Markers of kidney damage such as the albumin:creatinine ratio indicate early diabetic kidney disease. Improvement in proteinuria can be demonstrated up to 2 years postoperatively following bariatric surgery in patients with T2DM [79]. Bariatric surgery can improve albumin:creatinine ratios over a 5-year period [80]. Case–control data at 10 years comparing BPD and medical therapy show a steady decline in renal function in the medical group, with stabilization of eGFR and remission of albuminuria in the BPD group [81].

For retinopathy and neuropathy, there are fewer data on outcomes following bariatric surgery. Results from a cohort study 1 year after RYGB, VSG, and AGB report no change in retinopathy scores in a group with T2DM [70]. This is reassuring, given the dramatic improvements sometimes seen in glycemic control after bariatric surgery, which could be expected to worsen retinopathy [6,7,76].

Bariatric surgery does not have a proven role in treating those with microvascular complications, but early evidence suggests that it may be useful, at least in diabetic kidney disease [77,78,80,82]. There are no data to support the use of bariatric surgery in treating diabetic retinopathy or neuropathy, but the available data suggest that it is not deleterious in retinopathy. Given that there are no micronutrient or vitamin deficiencies, then it would not be expected to worsen neuropathy.

**Mechanisms of diabetes remission and improvement in glucose homeostasis**

There are several postulated mechanisms responsible for the improvement in glucose homeostasis after bariatric surgery. It is widely recognized that weight loss has major positive implications for metabolic health, and therefore it has been thought that the metabolic benefits derived from bariatric surgery are simply due to weight loss alone. However, there are multiple other mechanisms that can additively improve metabolic outcomes after bariatric surgery. They include hormonal changes, effects on appetite, and changes in insulin sensitivity. In fact, these all probably act in concert, but each mechanism will be discussed individually.

**Insulin resistance**

Insulin resistance is associated with obesity, and thus is a key factor in the development of most cases of T2DM [10–15]. Hepatic insulin resistance manifests as a rate of glucose production that is inappropriately elevated for the prevailing plasma glucose and insulin levels. Peripheral insulin resistance manifests as impaired insulin-mediated glucose disposal by target tissues (mostly, skeletal muscle and adipose tissue). The latter is directly measured by the euglycemic hyperinsulinemic clamp technique, while the former requires the use of glucose tracers (HOMA-IR being a crude surrogate). Hepatic insulin resistance...
is present in obese subjects with or without T2DM. Calorie restriction lowers plasma glucose and insulin concentrations and hepatic glucose production, whereby hepatic insulin resistance is improved. As early as a few days after surgery, before any weight loss, hepatic insulin sensitivity ameliorates likely due to the severe caloric deficit, with reductions in resting metabolic rates [87–89]. These improvements continue for up to 12 months postoperatively, and have been shown in both animal and human studies [90–92].

With regards to mechanisms, the “foregut hypothesis” posits that bypass of the proximal small bowel reduces the secretion of unknown gastrointestinal factors that decrease insulin secretion and promote insulin resistance. Accordingly, duodenal exclusion would reduce production of these putative “anti-incretins” leading to an increase in insulin secretion and action [93]. However, clamp studies have consistently shown that peripheral insulin resistance does not improve after RYGB as promptly as hepatic insulin resistance [87,91,94,95]. In the longer term (up to 6 months), insulin-mediated glucose uptake is enhanced by RYGB in both diabetic and nondiabetic recipients, and quantitatively the improvement is roughly proportional to the amount of weight loss [95]. Furthermore, in studies comparing RYGB and AGB with equivalent weight loss, clamp-based insulin sensitivity was found to be similarly increased, confirming that in humans the bypass itself does not exert a weight-independent effect on insulin action [96]. In contrast, both BPD and ileal transposition have been shown to produce a rapid and partially weight-independent enhancement of insulin action in subjects both with and without T2DM [97–99]. This has led to a “hindgut hypothesis” whereby anticipated nutrient delivery to the distal gut triggers responses (hormonal signals, substrates, neural reflexes) that potentiate insulin action [100,101].

In summary, bariatric surgery impacts on insulin action through different mechanisms and in different time courses. In the early postoperative period, calorie restriction improves hepatic insulin resistance, and later, weight loss restores peripheral insulin sensitivity. Throughout the postoperative period, the fall in glycemia relieves glucose toxicity. Altering the sequence of nutrient contact with the distal gut mucosa likely engages additional mechanisms of insulin sensitization. This does not fully account for the effects of bariatric surgery on glucose homeostasis, and enhanced hormonal activities postoperatively, including insulin secretion, are major contributors.

**Insulin secretion and β-cell function**

β-Cell function is an important component of the effect of bariatric surgery on glucose metabolism, especially in the diabetic patient, and impacts β-cell function by improving secretory dynamics in the face of reduced absolute insulin secretion rates. In fact, both the acute insulin response to intravenous glucose (AIR) and glucose sensitivity in response to oral glucose or mixed meals are improved early after surgery, and continue to improve as insulin resistance abates. Very low-calorie diets are associated with improvements in first-phase insulin secretion in obese cohorts with diabetes as compared to BMI-matched control groups [102]. While this may be a contributing factor, incretin hormones have been implicated in the improvements in insulin secretion after surgery, and GLP-1 is consistently associated with improved insulin secretion [103,104]. Bariatric procedures vary in how they enhance the incretin response, and RYGB has a greater effect on these hormones than AGB [5,105]. RYGB results in increased insulin secretion after eating when compared to patients with equivalent calorie restriction over 2 weeks in nondiabetic matched groups with matched weight loss [95]. An enhanced postprandial GLP-1 response is associated with this improved insulin secretion [88,90,94,95].

Here, the foregut and hindgut hypotheses may converge in that an accelerated delivery of nutrients to the distal gut improves β-cell function through an enhanced secretion of gut hormones such as GLP-1 and PYY. Thus, both VSG and RYGB potentiate β-cell function from the early postoperative period [106]. Support for this general mechanism comes from studies examining the effects of flexible plastic sleeves that prevent nutrient contact in the duodenum. These devices result in improvement in glucose homeostasis shortly after insertion [107]. Further support comes from experiments involving ileal interposition [104,108]. In this operation, a segment of the ileum rich in GLP-1 and PYY secreting cells is transplanted into the upper intestine, near the duodenum-jejunum boundary, thereby increasing exposure to ingested nutrients [108].

It must be emphasized that in T2DM, bariatric surgery may improve β-cell function but may fail to restore it to the level of nondiabetic individuals. Correspondingly, diabetes remission may be incomplete or transient. Available evidence indicates that the strongest predictor of failure to induce remission of diabetes is low β-cell function preoperatively, such as is seen in patients with long-standing severe hyperglycemia [109].

**Appetite and energy expenditure**

There is no doubt that reduced food intake modulates improvements in glucose homeostasis after bariatric surgery. Bariatric surgery is associated with improved satiety and reduced appetite through central and humoral mechanisms [32,75,110]. This effect is regulated through a complex neuroendocrine network that is still being defined [110,111]. RYGB can increase the secretion of satiety promoting gut hormones including PYY and GLP-1 [112,113]. The neural pathways regulating energy intake and energy expenditure includes the arcuate nucleus, which has PYY receptors [111,114]. Pharmacologic antagonism of PYY and GLP-1 results in increased appetite in recipients of RYGB and AGB [112]. Therefore, the enhanced gut hormone response has a central effect on appetite after surgery [32,114].

However, there are also effects on enteronuclear systems involving the vagus nerve and its network. The contribution of GLP-1-mediated vagal responses in regulating appetite is not fully understood [115,116]. Selective vagotomy can impair glycemic control, and have some effect on appetite, but vagotomized rats eat less than controls [116]. Studies comparing
central and intraperitoneal administration of GLP-1 in vagotomized rats suggest that there is significant contribution to satiety from both central and vagal incretin stimulation [111]. Further studies are needed to establish how the enteroneocrin system interacts with the central nervous system to affect satiety.

Changes in the pattern of substrate oxidation may affect the appetite/satiety balance. During surgically induced weight loss, there is preferential oxidation of fat over carbohydrate due to lower plasma insulin concentrations and enhanced lipolysis [87,117]. Raised circulating ketone bodies characterize the phase of rapid weight loss postoperatively [118]. GLP-1 and the GLP-1 receptor agonist, oxyntomodulin have also been implicated in the postoperative decrement in calorie intake, whether by direct central relays or through the slowing of gastric emptying [119–122].

**Mechanisms of diabetes remission: other perspectives**

Alterations in the total or individual species levels of bile acids in the gut and circulation have been implicated in improvements in glucose metabolism induced by bariatric surgery. Bile acid levels in the plasma are increased after RYGB, which has a significant early effect on improving glucose homeostasis, whereas AGB does not change plasma bile acid levels [123–126]. The level of bile acid fractions in plasma negatively correlates with glycemic excursions, implicating bile acids as agents in glucose homeostasis [123].

Bile acids can directly or indirectly affect glycemic control through several pathways, including the TGR5 receptors or nuclear FXR receptors, and the release of fibroblast growth factors (e.g. fibroblast growth factors 19 and 21), which exert actions on a wide range of tissues including the hypothalamus [127–129]. A central effect on food intake and appetite is possible, as bile acids can cross the blood–brain barrier and act on receptors in the hypothalamus [127–131].

The effect of bile acids may be to augment postprandial secretion of GLP-1 and other gut hormones [124]. Increased bile acid secretion after RYG demonstrates association with greater gut hormone levels [41]. Plasma bile acids are also elevated in animal models of VSG [132]. Bile acids could affect food intake, energy expenditure, and glycemic control through their actions on membrane TGR5 receptors or nuclear FXR receptors and the release of fibroblast growth factors. However, in the absence of detailed mechanistic studies, the exact role of bile acid as mediators of weight loss and glycemic control after RYG remains unclear. Preclinical and clinical studies are underway to determine their physiologic effects, and their therapeutic potential.

Gut microbiota in the context of obesity and weight loss have also been identified as important metabolic mediators after bariatric surgery, and bacteria have been implicated in the development of obesity [133]. Bariatric surgery can alter the gut microbiota, and a depletion of *Prevotellaceae, Archea, Firmicutes, Bacteroidetes* colonies, and an increase in the *Bacteroidetes/Prevotella* ratio and Gammaproteobacteria of this flora has been observed after RYGB [134–136]. These alterations can be associated with weight loss. The etiology of these changes is unclear, but may involve alterations in dietary macronutrient composition, anatomical manipulations, pH, and bile flow.

While the microbiota may change as a result of surgery, they have been shown to affect the surgical recipient. Adoptive transfer of gut bacteria from RYGB recipient mice to unoperated germ-free mice results in weight loss and increased energy expenditure [137]. This novel finding suggests that probiotic therapy could modulate energy homeostasis and potentially glycemic control. However, the exact mechanisms through which gut bacteria contribute to weight loss remain to be determined, and similar studies remain to be completed in humans.

**Conclusion**

Bariatric surgery is an effective treatment when combined with best medical care for obesity and obesity-related comorbidities including diabetes. However, many questions remain unanswered, including the role of bariatric surgery in treating metabolic diseases beyond diabetes, and the effect on diabetic complications.

The mechanisms of the effects of bariatric surgery are multiple, and include changes in body weight, appetite, food intake, gut hormone secretion, and resultant improvements in insulin resistance and insulin secretion. The effects on energy expenditure may also be important. Other mechanisms such as bile acids and microbiota are likely to also play a role, but more data on these pathways are needed.

**References**

Chapter 34
Obesity Surgery

gica* 1976;31(10): 560–566. [Italian]
95 Campos GM, Rabl C, Peeva S, et al.: Improvement in peripheral glucose uptake after gastric bypass surgery is observed only after substantial weight loss has occurred and correlates with the magnitude of weight lost. *Journal of Gastrointestinal Surgery* 2010;14(1):15–23.


Animal models of obesity and type 2 diabetes

David A. York, Mie Jung Park-York and Timothy Gilbertson
Department of Biology, Utah State University, Logan, UT, USA

Key points

- Application of new genetic technologies has facilitated the development of many new models.
- Genetic technologies have enabled proof of function experiments for specific genes or specific signaling pathways.
- The functional role of a specific protein in a specific cell type can now be studied by targeted gene knockouts or expression.
- Simple organisms (C. Elegans or Drosophila) also provide useful models for studying specific gene functions.
- Epigenetic changes associated with the early in utero and neonatal environment may have major effects on the development of obesity and its associated comorbidities.
- Recent identification of the role of the gut microbiome in promoting health has led to the development of numerous animal models that have enabled greater insight into potential mechanisms.
- Animals provide good models for understanding the mechanisms through which bariatric surgery has its beneficial effects in reducing obesity and type 2 diabetes.
- Application of animal modeling approaches will help to identify the physiologic roles of fatty acid binding proteins.

Introduction

Animal models have made a huge contribution to our understanding of obesity and type 2 diabetes (T2DM), providing insight into the causes and the pathophysiology of the conditions and their molecular and genomic underpinnings as well as providing evidence for new potential therapeutic targets [1,2]. From all of these studies has come the recognition that obesity and T2DM results from an interaction between the organism’s genetic susceptibility and the environment in which it lives. This is true of both vertebrate and nonvertebrate animals. Thus it is not surprising that there is now a voluminous literature on environmental and genetic influences that can modulate the susceptibility to becoming obese and diabetic. However, with the development of new and selective targeting techniques, the focus has changed from identifying genes that cause, predispose, or facilitate the development of obesity and T2DM to one in which selective gene targeting is used to develop insight into the roles of specific signaling pathways and for proof of function of specific genes. It is not the intention of this chapter to provide a comprehensive review of all these animal models. Indeed this would take a book by itself. Readers are referred back to a previous review for information on earlier models [1]. Instead, the focus will be on identifying the differing approaches that have been and are being utilized in animal studies more recently to further our insight into these metabolic conditions, while providing some specific examples of the insight that has been gained. In addition, we shall focus on two emerging areas where the use of these technological approaches is likely to lead to real advances in our understanding in forthcoming years.

Early models of diabetes and obesity were mainly in spontaneous mutations identified in laboratory rodents that were subsequently identified as being related to leptin signaling or melanocortin signaling or in studies across rodent strains that significantly differed in their response to high-fat diets (HFD) [3,4]. While the ability to manipulate gene expression in the whole animal has effectively maintained the focus on mice and rats, there has also been an explosion of interest in other species from the nematode worm Caenorhabditis elegans (C. elegans) and the Drosophila fruit fly up to primates. The latter have provided substantial insight into the temporal changes associated with the development of T2DM in obesity [5]. Since nutrient sensing is a requirement for all levels of organism, studies in simple organisms has helped to identify genes that may function in metabolic control in higher organisms. Thus flies with an insertion at the Atf4 locus lack fat and have increased susceptibility to starvation. Subsequent knockout of the Atf4 gene in mice identified a similar function; the mice were lean and were resistant to diet-induced obesity, diabetes, and hyperlipidemia [6]. The nematode C. elegans provides an inexpensive model that has attracted increasing attention for
a wide range of studies as it includes many of the same genes or gene homologs that are found in higher animals. Its short lifespan, rapid accumulation of fat and known genome sequence have provided an excellent model for studies including genetics, feeding and satiety, metabolism, as well as the screening of potential therapeutics. For example, study of the tub-1 mutation in C. elegans has provided insight into the pathways through which this gene regulates fat storage and lifespan [7].

Environmental influences
Dietary composition, particularly the level and type of dietary fat and carbohydrate is the most common environmental change that is associated with the development of obesity in a wide range of animal species and humans [8]. Significant strain differences in the susceptibility to become obese and/or diabetic on a HFD have been identified in mice and rats [9], a clear indication of the multigenic genetic variability in susceptibility. Indeed, the response to introduction of a HFD has become the classic experimental condition that is applied to phenotype all genetically manipulated animals for their susceptibility to become obese. Numerous gene linkage studies have been applied to these strain differences with only limited success in identifying genes of causality. However, selective breeding of animals prone to or resistant to obesity has enabled the identification of numerous physiologic differences associated with these different susceptibilities. An example of these are the studies of Levin and colleagues [10,11] which have identified differences in CNS signaling that could be linked to the differing susceptibility to become obese. These include differences in axonal pathways between the arcuate and PVN in the hypothalamus that are associated with both leptin and insulin regulation of feeding behavior through actions on the NPY/AgRP and POMC neurons [11]. Notwithstanding this, manipulation of the fatty acid composition of HFD has identified the differential effects of saturated as opposed to N-3 polyunsaturated fats in promoting or protecting, respectively, against the development of diabetes. Other dietary manipulations that have provided insight into the development of obesity include high sucrose, cafeteria diets, and diets with excessive zinc content. Alternatively, even the provision of standard laboratory chow can induce obesity and diabetes in susceptible rodent strains, for example the sand rat whose normal desert habitat has a scarcity of food supply.

Early environment and epigenetics
In recent years there has been the exciting recognition that the environmental influences on the development of obesity and diabetes are far greater than diet alone. From the early studies of Barker and colleagues [12] on the effects of maternal malnutrition on the health of the adult offspring, many animal models for the epigenetic regulation of obesity and T2DM have evolved. Studies in rodent species as well as larger animals have shown the association between maternal obesity and a sarcopenic form of adult obesity in the offspring who develop a metabolic syndrome phenotype associated with increased fat and reduced muscle mass, insulin resistance, and hypertension [13,14]. Further studies of this model suggest that the chronic inflammation that is associated with maternal obesity may expose the fetus to increased proinflammatory cytokines and reactive oxygen species [15]. These changes may persist across generations through both maternal and paternal lineages.

Obese mothers also potentially expose their offspring to higher levels of insulin and leptin during both fetal development in utero and during the early suckling period. Insulin injections into the dam or direct injections into the hypothalamus during the first 8 days of life are both associated with the development of future obesity in the adult. It is also clear that leptin injections into neonatal ob/ob mice can repair the defects in axonal pathways between the arcuate and PVN. Hence, it is not clear if the phenotypic results of these early life manipulations represent epigenetic changes or changes in the development of neuroanatomical pathways and functions.

The studies of epigenetic effects have not been limited to rodent models but has also been investigated in farm animals such as sheep and pigs. These epigenetic changes in gene expression have been associated with modification of histones by acetylation and methylation and with methylation of the promoter regions of genes. For example, moderate maternal undernutrition of sheep was associated with changes in histone methylation and acetylation of the glucocorticoid receptor (GR) gene whereas promoter methylation led to changes in POMC gene expression in the offspring [16]. Similarly maternal HFD may significantly change the expression of genes in multiple brain regions that lead to alterations in behavior and energy balance [17].

Viral infections
Viral infections have long been associated with the development of obesity through their potential effects on hypothalamic regulatory systems [1] but more recently the focus has been on adipogenic viruses such as Adenovirus Ad36. This has been linked to mild obesity in a wide range of animal strains including humans. Ad36 is thought to upregulate adipocyte differentiation and proliferation as well as enhance triglyceride synthesis and deposition [18] but, at the same time, it improves glucose homeostasis. This is thought to be the combined result of Ad36 inhibition of IRS-1 tyrosine phosphorylation that reduces insulin-stimulated glucose transport in the presence of Ad36 activation of RAS signaling that independently enhances PI3K signaling and glucose transport [19].

The gut microbiome
Another major environmental influence on the development of obesity and T2DM is the gut microbiome [20]. Immediately after birth, the gut becomes rapidly populated with microorganisms that develop a symbiotic relationship with the host.
Germ-free mice are much leaner than conventionally raised mice [21, 22] possibly due to the release of a lipoprotein lipase inhibitor and activation of hepatic and muscle AMP kinase that favors fatty acid oxidation. This may in turn reflect the absence of short chain fatty acids or immune activators such as bacterial lipopolysaccharide (LPS) normally produced by the microbiota in germ-free mice. High-fat diets lead to an increase in gram negative bacteria that produce LPS and this in turn activates proinflammatory cytokine production that will promote obesity and T2DM. Thus comparison of germ-free and conventional mice becomes an important model for demonstrating the role of gut microbiota on host metabolism and behavior [23]. Diet-induced obesity has been linked to major changes in the gut microbiome that are reversible by weight loss [24] and similar changes have been identified in humans. Prebiotic and antibiotic therapies also provide models for studying the role of the microbiota as do mice with mutations in the MYD88 gene that is associated with the toll-like receptor response. Studies in these models indicated a role for the cannabinoid 1 receptor in modulating gut permeability [25]. Likewise, mice that lack toll-like 5 receptor in the gut mucosa are less able to defend against bacterial infection and develop a metabolic syndrome type phenotype that includes increased body weight and body fat, insulin resistance, hyperlipidemia, and hypertension [26]. Transfer of the gut microbiota from such mice into germ-free wild-type mice induced many of the metabolic syndrome characteristics suggesting the importance of the gut microbiota to the induction of metabolic disease [26]. Future studies of the role of the human gut microbiome may be aided by the transfer of the human microbiome into pigs [27].

Animal models, other than those that arise spontaneously, have generally been developed to explore the functions of specific genes or signaling pathways or to develop models for human disease. One exception to this is the microbiome models described earlier. Another example arises from the success of bariatric surgical approaches to reverse gross obesity and its comorbidities in human populations [28, 29]. The mechanisms through which these various surgical approaches lead to a rapid improvement in comorbidities, such as T2DM, that appear to be independent of the weight loss, are not really understood. Changes in gastrointestinal anatomy have long been known to have significant impacts on energy balance [30]. Since gastrointestinal–brain communication is particularly difficult to investigate in humans, this has promoted the development of a number of animal models in which the gastrointestinal architecture has been altered to mimic those surgical procedures used in humans. These models have focused on sleeve gastrectomy, Roux-en-Y gastric bypass, and ileal transposition [31–33]. They allow detailed investigations of the temporal changes in gut and brain physiology that may identify the mechanisms that underlie the beneficial effects of these surgical procedures and when performed in genetically manipulated animals [33] can identify the potential role of specific signaling systems to the surgical response.

Genetic models

The initial studies of the genetic basis for obesity and T2DM relied upon spontaneously arising mutations in mouse and rat models. Study of such models led to the identification of the leptin gene (ob/ob mouse), the leptin receptor (db/db mouse and fa/fa and Koletsky rats), agouti and the melanocortin signaling system (A/y/a mouse) and did much to identify the signaling pathways in the CNS through which leptin and insulin can act to modulate energy balance. These have been extensively reviewed [1, 34, 35]. However, the subsequent development and use of a wide range of gene targeting techniques has led to a plethora of animal models. Thus, we have moved from these natural mutations through whole animal transgenic expression or gene knockout to techniques that allow tissue-targeted knockouts and overexpression and now, on to gene manipulation in specific target cells (Table 35.1). Such models have not only allowed the “proof of function” type of experiment but also allow detailed insight into the signaling pathways within the CNS and peripheral tissues that promote obesity and diabetes. In addition, they have enabled definition of the neuronal phenotypes and anatomical pathways through which energy balance and peripheral metabolism are regulated. Some examples of these approaches are provided in the following sections.

The identification of leptin and its signaling receptor from studies of natural mutation mice and rat models provided a major step forward in our understanding of the systems that underlie the regulation of energy balance. It identified leptin as a cytokine produced in adipose tissue that acted as a feedback signal to regulate food intake and energy balance [36]. This focused studies onto the hypothalamus where leptin receptors are expressed in both AgRP/NPY and POMC/CART neurons within the arcuate nucleus, neurons known to exert major stimulatory and inhibitory functions on food intake while also having anabolic and catabolic effects respectively on peripheral metabolism. Many of these effects were identified through animal models in which specific genes (AgRP, NPY, POMC, MC4R, and MC3R) were knocked out. This led to the general hypothesis that leptin inhibited the AgRP/NPY neurons and activated POMC/CART neurons to suppress feeding, activate sympathetic drive to the periphery and enhance catabolic pathways. However, the ability to target gene knockouts to specific cell types using Cre-Lox technology has enabled significant new insight. This has been accomplished in several ways, either by breeding the Cre and Lox animals together to knockout a specific gene universally or in specific cell types or in specific cell types within a local area by, for example, stereotaxic injections of adeno- or lenti-viral vectors containing the Cre recombinase (Figure 35.1). These approaches have provided new and unexpected insight into these regulatory pathways. For example, rescue of leptin receptors in the arcuate nucleus of db/db mice only partially reversed their hyperphagia [37] while the use of Cre/Lox technology to specifically delete leptin receptors in POMC neurons [38–40] only produced mild
Table 35.1 Progression of genetic approaches used in animal models

<table>
<thead>
<tr>
<th>Approach</th>
<th>Example</th>
<th>Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Spontaneous models of genetic obesity</td>
<td>Single gene mutations (Lep, Lepr, Agouti)</td>
<td>Obesity ± Diabetes</td>
</tr>
<tr>
<td>2. Whole animal transgenics (Tg/KO)</td>
<td>Tg AgRP with β-Actin promoter MC3R KO</td>
<td>Obesity and coat color change</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Obesity without hyperphagia</td>
</tr>
<tr>
<td>3. Tissue targeted Tg or KO</td>
<td>Hepatic insulin receptor (Insr) KO</td>
<td>Diabetes</td>
</tr>
<tr>
<td></td>
<td>Hsd11b1 Tg on aP2 promoter</td>
<td>Central adiposity</td>
</tr>
<tr>
<td>4. Gene manipulation in specific cell type</td>
<td>Lepr KO in POMC neurons</td>
<td>Mild obesity</td>
</tr>
<tr>
<td>5. Gene rescue in KO mice</td>
<td>MC4R rescue in PVN of MC4R KO mice</td>
<td>Prevented 50% of obesity</td>
</tr>
<tr>
<td>6. Expression of a designer receptor in specific neuronal cells (DREADD)</td>
<td>Muscarinic receptor in POMC neurons</td>
<td>Feeding when receptor activated by ligand</td>
</tr>
</tbody>
</table>

Further information on each of these examples can be found either in the text or in Reference [1].

Obesity, indicating that these neurons only partially regulate the full effects of leptin on energy balance and suggest that leptin receptor activity in other brain regions is essential for the leptin effects on feeding behavior and metabolic rate. In contrast, leptin activation of POMC neurons is necessary for its ability to reduce glucose and insulin levels and to raise blood pressure [40].

Insulin, besides its role to control peripheral glucose homeostasis, also acts as a chronic feedback signal to the CNS on the level of body fat and energy stores, as does leptin. Such a role for insulin was initially promoted from ICV infusion studies which showed insulin could suppress food intake. However, it was the ability to knock out the insulin receptor in a tissue-specific manner that provided the conclusive evidence of this long-term role of insulin as a feedback signal regulating energy homeostasis. Mice with a neuronal specific loss of insulin receptors are hyperphagic and become moderately obese and have an attenuated ability to suppress hepatic gluconeogenesis [41]. This showed the importance of central insulin signaling to the regulation of peripheral metabolism. However, knockout studies can always be criticized for the potential changes that occur during development in the absence of the expressed gene, so the development of conditional knockouts or gene expression models has facilitated a clearer insight into the roles of specific genes in the growing or adult animals. One such example of this approach [42] compared two inducible knockouts of the insulin receptor that used the Cre-Lox system to produce mice that could knockdown IR expression through tamoxifen or doxycyclin treatment in either the whole body or just in peripheral tissues. By comparison of their responses, insight into the role of brain IRs was gained. These studies confirmed the importance of brain IR and central insulin signaling for the regulation of peripheral glucose homeostasis. Thus the roles of insulin to inhibit hepatic gluconeogenesis and glycogenolysis, to promote lipid deposition in adipose tissue and to act as a CNS signal to regulate energy homeostasis and peripheral metabolism are confirmed through these models.

Further support comes from studies of experimental models that have altered brain insulin sensitivity. This can be illustrated by the studies on PKCθ, a serine kinase that prevents activation of insulin receptor substrate 1 (IRS-1) through phosphorylation and promotes internalization [43,44] and degradation of the insulin receptor [45]. The role of PKCθ in regulating energy balance and insulin sensitivity, however, is unclear at this time, as shown by numerous conflicting reports. PKCθ null mice are protected against the insulin resistance that is induced by acute infusions of lipids [46], and knockdown of arcuate PKCθ by injections of PKCθ shRNA reduces weight gain on a HFD [47]. However, PKCθ is required for the maintenance of muscle insulin sensitivity [48]. Likewise, our own studies on PKCθ null mice [49] identified an increased susceptibility to obesity and insulin resistance consistent with a previous report of the effects of transgenic muscle expression of a dominant negative PKCθ [48]. The differing observations may relate to acute versus chronic responses, the level of obesity, diet differences, or tissue-specific effects of PKCθ. In contrast, hypothalamic PKCδ has been shown to regulate hepatic glucose production and mediate the effects of peripheral lipid signals [50].

The role of PKCδ in the brain has received little attention. It is widely expressed within the CNS albeit at low levels [47,51] as PKCδ is the major isof orm of PKC in neural tissues. Hypothalamic PKCδ has been shown to regulate hepatic glucose production and mediate the effects of peripheral lipid signals [50]. Nevertheless, knockdown of PKCδ in the arcuate nucleus with shRNA improved glucose tolerance in mice fed a HFD [47]. In our own studies, we have shown that PKCδ overexpression, induced by stereotaxic injections of a lentiviral-PKCδ construct into the central nucleus of the amygdala, induces
insulin resistance in that region. Chronically, this is associated with a small increase in food intake, increased body weight and body fat, the development of hepatic insulin resistance and increased hepatic triglyceride levels while conversely, the hypoglycemic response to insulin is increased and the glucose tolerance is improved suggestive of increased peripheral sensitivity [52,53]. This is a distinctly different phenotype from that induced by hypothalamic insulin resistance. Insulin acts in the hypothalamus to inhibit hepatic glucose production [54] and fatty acid activation of hypothalamic PKCθ suppresses this effect [47]. Intracerebroventricular (ICV) injections of PKCθ, which increased the expression of PKCθ in the arcuate nucleus, impaired glucose tolerance in our studies when rats are fed a HFD [52,53] consistent with a previous report that arcuate injections of PKCθ shRNA improved glucose tolerance [47]. This suggests that developing insulin resistance in the hypothalamus is associated with peripheral insulin resistance and/or increased hepatic glucose production whereas insulin resistance in the CeA is associated with improved glucose clearance. The lipid phenotype also differs between these two models with hepatic hypertriglyceridemia and normal serum triglycerides with CeA insulin resistance and normal liver lipids but increased serum triglycerides associated with hypothalamic insulin resistance.

Since insulin receptors are widely distributed in the CNS, we should be careful not to assume that all of the metabolic changes associated with central insulin activities are reflective of actions solely in the hypothalamus. The suggested role of the amygdala in regulating peripheral glucose and lipid metabolism is supported by recent reports that the medial amygdala contains glucose-sensing neurons that appear to innervate the VMH through Urocortin 3 neuronal connections [55].

**Tissue-targeted knockout or knockdown of insulin receptors**

Tissue-targeted knockout or knockdown of insulin receptors has not only provided profound new insight into the regulation of peripheral metabolism but has identified the importance of tissue cross-talk. This tissue-targeted gene knockout is achieved by using tissue-specific gene promoters to regulate Cre expression (Figure 35.1). Targeting hepatic insulin receptors using the albumin promoter to drive Cre expression led to a diabetes model with increased gluconeogenesis and hepatic insulin resistance and very elevated insulin levels that result from reduced hepatic clearance of insulin and increased secretion. Insulin normally activates multiple signaling pathways within the liver. Conventionally, it would have been necessary to have selective specific inhibitors of these individual signaling pathways to identify the specific roles of each. However, we now use genetic targeting for studies which will identify the role of specific signaling pathways. Thus knockout of FOXO1 reversed most of the metabolic changes associated with liver specific knockout of INSR, IRS-1, or IRS-2 genes. Such data was used to support the concept that insulin stimulated Akt signaling inhibited gluconeogenesis through suppression of FOXO1 gene expression [56]. This has been investigated further by Lu and colleagues [57] using double knockouts of Akt1/Akt2, and FOXO1 providing additional insight into the regulation of hepatic metabolism during fasting and refeeding.

Tissue-targeted knockouts of the insulin receptor have been particularly informative. These have been reviewed previously in detail [58]. Loss of adipose tissue (both white and brown) insulin receptors led to a lean mouse that was resistant to the development of obesity from dietary or hypothalamic manipulations while maintaining normal glucose tolerance.
suggesting that adipose tissue is not a major site for glucose clearance. In contrast, specific knockout of insulin receptors in brown adipose tissue using the UCP1 driven Cre expression leads to an age-dependent loss of BAT and, surprisingly, hyperglycemia associated with a loss of β-cell function. Loss of insulin receptors in muscle using the Cre/lox system impaired insulin signaling and insulin-stimulated glucose uptake but did not result in whole body insulin resistance. However, the mice do become obese and hypertriglyceridemic as more glucose is partitioned to adipose tissue for lipogenesis. In contrast, targeted lifetime knockout of IR in the pancreatic β cells inhibits glucose-stimulated insulin secretion that was not present in mice with a conditional knockout in the adult animal, indicating the developmental importance of insulin signaling in the pancreas.

Tissue-specific gene manipulations have also illustrated the interdependence and cross-talk between tissues. Thus, an adipose tissue specific knockout of peroxisome proliferator-activated receptor gamma (PPARγ) impairs adipocyte proliferation and increases free fatty acid flux to the liver which becomes insulin resistant and exhibits enhanced gluconeogenesis and triglyceride deposition [59]. In contrast, transgenic expression of a dominant negative cAMP response element binding protein (CREBP) in adipose tissue protects the liver from the development of steatosis on a HFD and enhances insulin sensitivity during the development of obesity [60].

While gene knockout provides many informative models, targeted gene rescue in knockout mice has become a frequent approach towards “proof of function.” An example of this approach by Balthasar et al. demonstrated that rescuing MC4R expression in the PVN of LoxTB-MC4R mice (which have deficiency of MC4R everywhere in the body) prevented 50–60% of the usual obesity associated with MC4R deficiency, and this was mostly mediated by a reduction in food intake since no effect on thermogenesis was observed [61]. Thus, MC4R activity in other brain regions regulates the melanocortin control of thermogenesis.

Finally, the “designer receptors exclusively activated by designer drugs (DREADD) technology” has been developed to facilitate the activation of specific neuronal subtypes. A mutated muscarinic G-protein coupled receptor, that is unable to bind native ligands but has a high potency for an otherwise inert pharmacologic ligand, can be targeted to specific neuronal subtypes expressing Cre recombinase using an adenoviral (AAV)-Cre-recombinase-dependent expression system that can be stereotaxically injected. Using this technology, Krashes et al. [62] injected an AAV-Cre recombinase expression-dependent mutated receptor vector into the arcuate nuclei of mice expressing Cre-recombinase in their AgRP neurons. Activation of these expressed receptors by injections of the pharmacologic ligand depolarized and activated the AgRP neurons and induced feeding behavior in a rapid and reversible manner. This DREADD approach will provide a series of new models that will facilitate the investigation of specific neuronal subtypes on both feeding behavior and peripheral metabolism.

### Fatty acid binding proteins

Given the extreme rise in obesity over the past two decades there has been a significant increase in the elucidation of the pathways used by the body to recognize and respond to nutrients at sites from the oral cavity through the digestive and central nervous systems. Of particular recent interest has been the identification of the mechanisms that the body uses to detect dietary fat. The recent deorphanization of a number of G protein coupled receptors (GPCRs) has lent credence to the notion that free fatty acids appear to be important primary chemical messengers that signal the presence of dietary fat but the physiologic function of many of these receptors has not been clearly defined at this time. The major classes of free fatty acids have been identified as ligands for these receptors including the mono- and poly-unsaturated fatty acids (GPR120 [63] and GPR40 [64]), the medium chain saturated fatty acids (GPR84 [65]) and the short chain saturated fatty acids (GPR41/43 [66]). In addition to the fatty acid-activated GPCRs, the fatty acid binding protein CD36 has been proposed to play major roles in the transduction pathways for fatty acids in a variety of fat-responsive tissues [67] either as a fatty acid transporter, primary receptor or co-receptor working in concert with fatty acid-activated GPCRs [68]. Together, it is clear that there is a picture emerging that free fatty acids, present in food or generated via the action of a number of lipases, signal the presence of dietary fat.

Research showing the ability of free fatty acids to act as ligands at GPCRs has spawned numerous studies aimed at identifying a role for these receptors in normal physiologic processes. These proteins vary in their distribution, ligand specificity, and functional roles as summarized in Table 35.2. GPR120 appears to play critical roles in the ability to elicit chemosensory responses in both the taste system and the digestive system to mediate the ability of fatty acids to stimulate the release of satiety hormones like glucagon-like peptide 1 and cholecystokinin from the small intestine [69], and to play a role in high fat diet-induced adipogenesis [70]. A closely related GPCR that is also expressed in the intestine, GPR119, whose cognate ligands include polyunsaturated alkamides, may also play a role in the control of food intake [71]. GPR40 expression overlaps with that of GPR120 and includes the taste system, gut, liver, pancreas, and adipose tissue and has been most well studied for its role in fatty acid-induced stimulation of insulin release from pancreatic β cells and in inflammatory responses [72].

GPCR-mediated physiologic responses to fatty acids are not limited to the unsaturated fatty acids. Additional members of the GPR40-family of receptors include GPR41 (FFAR2) and GPR43 (FFAR3), which have similar, but not identical, tissue distributions and functions [66] (Table 35.2). GPR41/43 respond to the short chain saturated fatty acids (SCFAs; C2:0 through C6:0) and like other fatty acid-activated GPCRs have roles in the enteric nervous system. Significant interest has focused on the reported effects of SCFAs generated through the activity of gut microbiota to stimulate these receptors in the digestive system related to a
number of functions including colonic contraction and intestinal anti-inflammatory responses [73] and in adipocytes where they contribute to SCFA-induced leptin release [74].

The fatty acid translocase, CD36, plays a critical role in the transport of long chain fatty acids across biologic membranes. While there a number of proteins in addition to CD36 that fulfill this function of fatty acid translocation in cells, CD36 appears to be of particular importance [75] for cell activation and the sensing of circulating fatty acid levels or fuel availability and to be the primary receptor for fat taste [76]. A link between CD36 function and obesity and metabolic syndrome [77–79] has been identified consistent with studies in human populations that appear to link variants in CD36 with differences in fatty acid sensing ability and, further, obesity [80].

In light of the described activities of these fatty acid receptive proteins, it is not surprising that a number of animal models have been developed to identify their causal link to nutrient-related diseases such as obesity and diabetes. With the exception of GPR84 where no direct link between activity at this receptor and obesity/diabetes currently exists, the other fatty acid receptive proteins have been implicated in these conditions. For example, the SCFA receptors, GPR41 and GPR43, which facilitate the role of SCFAs as endocrine regulators of metabolism may represent important targets for treatment of metabolic disorders (for review see [81]). Consistent with this interpretation, mice lacking GPR41 and GPR43 show reduced GLP-1 secretion and an impairment of glucose tolerance [82].

Like the fatty acid-activated GPCRs, by virtue of its function in fatty acid uptake, CD36 has similarly been shown to play important roles in metabolic regulation. Mice lacking CD36 have shown a host of physiologic challenges in a variety of tissues with many involving conditions related to metabolic syndrome and diabetes (for review, [83]). With respect to obesity and its health effects, it appears that the greater the expression of CD36 the greater the risk of obesity-related complications [84–86]. This viewpoint is strengthened by recent studies showing the regulation of CD36 expression by HFD [87,88] in a manner consistent with this relationship.

Our understanding of the nature of fatty acid signaling and the cognate receptors for fatty acids in the central nervous system and periphery is far from complete. However, a picture is emerging that shows important roles for fatty acid receptors in normal physiology particularly in processes related to the control of food choice, food intake, and in the control of energy balance and normal metabolic processes. The development of further animal models targeting these fatty acid signaling/transporting proteins will be fundamental to the development of further insight to their physiologic roles.

In summary, the development of new genomic technologies has dramatically expanded the number of new animal models of obesity and T2DM. These have promoted new insight into the pathophysiology of the diseases as well as providing additional insight into the roles of specific genes and signaling pathways. Their application to genes more recently implicated in lipid signaling will facilitate our understanding of these genes. Together with the development of models that mimic human conditions and treatment approaches, the continued use of and development of new animal models is likely to provide a fuller understanding of the roles of individual genes and signaling pathways in the pathophysiology of obesity and T2DM. This, in turn, may help in the future development of new treatment and prevention approaches.

References


53 Park-York M, Boghossian S, Oh H, York DA: PKCθ overexpression in the central nucleus of the amygdala or hypothalamus has differential effects on energy balance and peripheral glucose homeostasis. *Brain Research* 2013;1498:85 – 94.


69 Keller KL: Genetic influences on oral fat perception and preference: Presented at the symposium “The Taste for Fat: New Discoveries on the Role of Fat in Sensory Perception, Metabolism, Sensory Pleasure and Beyond” held at the Institute of Food Technologists.


CHAPTER 36
The role of the hypothalamus in the maintenance of energy balance and peripheral glucose control

Luis Varela¹ and Tamas L. Horvath²
¹Program in Integrative Cell Signaling and Neurobiology of Metabolism, Section of Comparative Medicine, Yale University School of Medicine, New Haven, CT, USA
²Yale Program in Integrative Cell Signaling and Neurobiology of Metabolism, New Haven, CT, USA

Key points
• The hypothalamic neuronal substrate of energy metabolism and glucose homeostasis is becoming clearer.
• Transcriptional regulators of AgRP and POMC neurons have been discovered and tied to leptin action.
• FOXO1 and mTOR are critical intracellular signaling molecules in hypothalamic regulation of feeding and glucose homeostasis.
• Different subpopulations of hypothalamic neurons respond to leptin and insulin.
• Regulation of feeding behavior and glucose metabolism do not always go hand-in-hand.

Introduction
Although obesity has only recently become one of the more important health problems of developed societies, it has been studied for centuries. Obesity is no longer looked at as just a major health problem but is considered an economic predicament as well—placing an enormous financial burden upon society for the care and treatment of patients. This metabolic state and its associated illnesses—such as diabetes mellitus, dislipidemia, and hypertension—have increased healthcare costs beyond sustainable levels. It has been calculated that over the next 20 years, the healthcare costs attributable to obesity will rise to about 16% of the total healthcare costs in Western countries [1,2].

Early attempts to understand the biologic and pathological aspects of obesity on the part of researchers have led to several observations. Several brain areas and factors are involved in the regulation of food intake and body weight as it relates to obesity. These studies found that the central nervous system (CNS), mainly the hypothalamus, plays a pivotal role in integrating signals from peripheral tissues—including the liver, white adipose tissue, gut, and so on—to regulate energy homeostasis (Figure 36.1). Secretions from these various tissues act as the peripheral signals that relay information regarding the energy status of the organism to the CNS, which integrates these signals and in turn regulates energy intake and expenditure. Two of the most well-known and studied peripheral signals are leptin and insulin, both of which are expressed in correlation with the total body fat mass [3]. In states of energy deficiency, the blood levels of these two molecules are lower, while in states of energy surplus, they are increased.

Earlier publications reported that pituitary adenomas cause obesity [4,5]; however, current reports using more innovative methods and novel techniques reveal that the hypothalamus is the major regulator of feeding. Efforts to understand the basis of obesity focused on it and its modulation by peripheral signals. The first clues about the importance of the hypothalamus in the control of feeding were given by different studies showing that hypothalamic lesions cause alterations in food intake and body weight [6–9]. The hypothalamus, located in the mediobasal part of the brain, is organized into well-structured nuclei (ventromedial (VMH), paraventricular (PVH), lateral (LHA), and so on). Within each nucleus are different neuronal populations characterized by the various neuropeptides that are expressed; these neuropeptides are involved in the control of food intake. For instance, it has been observed that lesions in the VMH are associated with hyperphagia [9], while lesions in the LHA cause hypophagia and a decrease in body weight [8]. Within the basal part of the hypothalamus is located the Arc, where the blood–brain barrier has a special modification to allow the entry of nutrients, hormones, and other molecules from the blood. Its privileged location allows it to be the first sensor of peripheral signals [10]. Because of this, it has been postulated that the Arc plays a fundamental role in sensing...
Leptin and insulin signaling pathways in the hypothalamus

Leptin, an adipokine, is secreted by white adipose tissue (WAT) in positive relation to the total amount of body fat in all species so far studied. Insulin is secreted by the pancreatic β cells, which are sensitive to the blood glucose level in the short term and the level of adiposity in the long term. Both peripheral signals can induce a strong anorexigenic effect [3]. Central administration of these molecules, which mimics a state of energy surplus, inhibits food intake and decreases body weight [12]. Furthermore, their central actions have a potent impact in regulating glucose homeostasis and lipid metabolism. The actions of leptin and insulin are mediated by the receptors LEPR-b and InsR, respectively. These receptors are widespread and are present in several areas of the CNS. Different studies have shown particularly significant expression in the hypothalamus, with higher expression in the AgRP and POMC neurons of the Arc. Given the role of these neuronal populations and the demonstration of Lepr and Insr within them, it was proposed some years ago that the leptin and insulin signaling pathways in the POMC and AgRP neurons of the Arc are highly significant in the regulation of food intake, body weight, and glucose homeostasis.

Leptin activates different signaling cascades assessed in culture as well as whole animal models (Figure 36.2). When this hormone binds the extracellular domain of LEPR-b, it recruits and activates the Janus kinase (JAK) pathway. JAK binds and phosphorylates LEPR-b [13], which activates STAT3 (Figure 36.2). Once phosphorylated, STAT3 binds to POMC and AgRP promoters, stimulating POMC expression and inhibiting AgRP expression [14,15]. In recent years, the most studied targets of leptin action are the proteins involved in the PI3K signaling pathway. Leptin activates PI3K, which induces the synthesis of PIP₃ from PIP₂ [16–18]. Accumulation of PIP₃ leads to PDK1 activation and thus, activates protein kinase B (PKB, also known as AKT) (Figure 36.2). Interestingly, the insulin signaling pathway converges with the leptin pathway as PI3K is activated by both to modulate body weight and glucose homeostasis [19,20]. AKT, one of PDK1’s downstream targets, plays an important role in the regulation and activation of a great number of proteins and transcription factors, including FOXO1 [21,22], AMPK [23,24], and mTOR [25]. FOXO1 acts as an inhibitor of expression of POMC [26]. When leptin and/or insulin phosphorylates this transcription factor, it induces its
The role of the hypothalamus in the maintenance of energy balance and peripheral glucose control

Figure 36.2 Leptin and insulin signaling pathways in the Arc. Signal transducer and activator of transcription 3 (STAT3) is activated and phosphorylated when leptin binds LepR-b. p-STAT3 binds to POMC and AGRP promoters, stimulating POMC expression and inhibiting AGRP. Leptin and insulin signaling pathways converge in phosphatidylinositol-3-kinase (PI3K). Activation of PI3K leads to activation and phosphorylation of Forkhead box protein O1 (FOXO1), a repressor of POMC expression. Phosphorylation of FOXO1 provokes its nuclear export and allows STAT3 binding to POMC and AGRP promoters.

nuclear export, thus allowing the binding of STAT3 to the POMC promoter to induce its expression. In AgRP neurons, the nuclear export of FOXO1 abrogates mRNA expression, because in this case, STAT3 is able to bind the AGRP promoter and impair transcription [27].

Leptin and insulin regulate the AMPK signaling pathway in the hypothalamus [24]. Proteins involved in this pathway sense the global energy status and are activated by an energy deficiency [23,28]. Leptin and insulin inhibit the activation of AMPK and its downstream targets, thereby signaling energy surplus. The mTOR pathway, closely related to AMPK, plays an important role in the hypothalamic regulation of food intake, sensing nutrient availability and the energy status [25,29]. In contrast to AMPK, mTOR expression is higher in situations of energy surplus [25]. Thus, it has been reported that intracerebroventricular (ICV) leptin administration increases mTOR expression [25]. Indeed, mTOR plays an important role as a mediator of leptin action in the hypothalamus: it has been demonstrated in rodents that an intact mTOR signaling pathway is necessary to maintain the anorexigenic effects of leptin [25].

**Hypothalamic leptin and insulin mediate food intake, energy expenditure, and glucose homeostasis**

**Leptin in POMC and AGRP neurons**

For a number of years, it has been known that leptin and insulin play important roles in the regulation of energy homeostasis [30]. In mice, mutations of leptin (Lep\(^{ob/ob}\)) or leptin receptor (LepR \(^{db/db}\)) are associated with increased body weight and impaired glucose homeostasis (hyperglycemia, hyperinsulinemia, insulin resistance, impaired hepatic glucose production, and so on) [31–33]. It has been reported that ICV leptin injections in these mice reduce body weight and food intake, and ameliorate glucose metabolism [12]. It is expected, then, that obese patients would present lower serum leptin levels, but surprisingly, the opposite is true—high levels are expressed. This is believed to be the consequence of the so-called leptin resistance, a state in which leptin signaling is impaired and the body compensates by producing excess leptin [34,35]. These findings, together with the evidence that the Arc neurons are very important in the regulation of energy balance, focused attention on the POMC and AGRP neurons as the major targets of leptin action. It has been reported that leptin action depolarizes (activates) POMC and hyperpolarizes (inhibits) AGRP neurons [36].

There were many attempts to investigate the exact mechanism by which these hormones act in the hypothalamus. Following the discovery and use of Cre/Lox technology, the mechanisms of action of both hormones have become better understood but are still unclear. The deletion or overexpression of LepR in POMC or AGRP neurons causes changes in body weight, food intake and glucose homeostasis [38–42]. Thus, the importance of POMC in leptin signaling has been intensely studied. Currently, it is well known that POMC neurons are essential in mediating leptin actions in the brain. Several animal studies have shown that deletion of LepR in these neurons promotes obesity without changes in food intake or energy expenditure.
and that overexpression of this receptor in POMC neurons of Lepr-null mice partially rescues the obesity and hyperphagic phenotype characteristic of this mouse strain [39,43]. Despite this, there are controversies about the role that POMC plays in leptin signaling. Two different reports have suggested two different reasons for the decrease in body weight seen in Lepr-null mice that re-express Lepr in POMC neurons. One report states that it is due to diminished food intake [39] while the other attributes it to increased energy expenditure [43]. In the latter, the authors postulated that this discordance could be due to differential Lepr expression in different subsets of POMC neurons, and that the effects of leptin outside of the Arc are important for the proper regulation of food intake. In both animal models, re-expression of Lepr improved glucose levels and insulin sensitivity independently of body weight, indicating that leptin signaling in POMC neurons plays a key role in regulating glucose homeostasis [42,43]. The effects of leptin on AgRP neurons, however, are not as clear. In these neurons, the deletion of Lepr is associated with an increase in total body weight and adiposity [44]. AgRP and POMC are both involved in the maintenance of energy balance by leptin signaling. Mice with ablation of Lepr in both the POMC and AgRP neurons present more severe obesity [40]. It has been postulated that these subsets of neurons work in a synergistic way to respond to leptin [40].

**Insulin in POMC and AgRP neurons**

Insulin, like leptin, has been demonstrated to be an important regulator of energy regulation via the CNS [45–48]. It has been demonstrated that central insulin administration regulates feeding behavior. Since the first evidence demonstrating that brain-specific deletion of InsR (NIRKO mice) is associated with obesity and increased food intake [45], a huge number of reports have tried to identify the neurons and mechanism through which this hormone affects the brain. In these NIRKO mice, obesity is related to high leptin levels [45], demonstrating that central leptin actions require intact insulin signaling in the CNS. Furthermore, central insulin is essential for the regulation of proper glucose homeostasis. As in the case of leptin, POMC and AgRP neurons are supposed to be critical targets of the insulin signaling pathway. It has been shown that insulin acts on these subsets of neurons, as evidenced by the fact that its central administration provokes hyperpolarization of POMC and AgRP neurons in the Arc [20,24,40,49]. Curiously, unlike leptin, the selective ablation of Insr in POMC or AgRP neurons has no effect on body weight or food intake [20]. Similar to previous pharmacologic reports, transgenic deletions of Insr in POMC and AgRP neurons have an impact on glucose metabolism. Specifically, similar to pharmacologic blockade of insulin signaling, mice with ablation of Insr in AgRP (but not in POMC) neurons have impaired suppression of hepatic glucose production (HGP) [20], demonstrating the significance of insulin action on these neurons in the maintenance of glucose homeostasis.

To further investigate the role of Arc neurons in mediating the effects of insulin, one work has analyzed the effects of Insr re-expression in AgRP and POMC neurons in mice lacking Insr expression in the Arc [50]. In accordance with previous reports, these Insr knockout mice showed a potent suppression in HGP when Insr is re-expressed in just the AgRP neurons [50]. In contrast, re-expression in the POMC neurons had no effect on HGP, but surprisingly, these mice have impaired glucose homeostasis compared to the Insr mice [50]. Contrary to other reports, POMC neurons seem to play a role in energy homeostasis through insulin signaling. POMC knockin mice show increased food intake and locomotor activity [50]. Taken together, these results demonstrate that AgRP neurons mediate many of insulin’s effects on glucose homeostasis and that other neuronal populations in different hypothalamic nuclei—such as VMH, for example—are also likely involved in the regulation of body weight and food intake.

**Leptin’s and insulin’s mechanisms of action in regulating energy and glucose homeostasis**

After the Arc POMC and AgRP neurons were identified as central mediators of leptin and insulin actions, several reports tried to understand the exact mechanisms through which these hormones modulate energy and glucose homeostasis. As stated earlier, leptin and insulin are involved in the activation of different signaling pathways. Genetic modifications in these target genes have provided some clues about how both hormones are able to trigger responses by the POMC and AgRP neurons. For instance, after STAT3 deletion in the CNS was found to induce an obese phenotype associated with lower mRNA levels of POMC, this gene was studied as an important factor involved in leptin signaling [51,52]. Various work in which STAT3 was deleted or overexpressed in Arc neurons demonstrated a clear role of this protein in this mechanism [14,15,53]. In POMC neurons, it has been found that STAT3 activation causes mild obesity associated with increased food intake and lower POMC mRNA levels [15], which is in agreement with the stimulatory effect of STAT3 on POMC expression mentioned earlier. Furthermore, this animal model exhibits an impaired leptin response. Peripheral leptin administration was unable to reproduce the expected effects on body weight and food intake seen in control mice, an observation that reinforces the importance of STAT3 signaling in POMC neurons regulating leptin effects [15,53]. In addition to leptin resistance, constitutive activation of STAT3 in POMC neurons leads to insulin resistance [15], which is probably mediated by an increased expression of suppressor of cytokine signaling 3 (SOCS3; an inhibitor of leptin signaling). This study confirms the role of STAT3 as a mediator of not only leptin, but also insulin signaling in POMC neurons.

The role of STAT3 in AgRP neurons has also been examined [14,26,54]. Its activation by leptin has inhibitory effects on...
AgRP mRNA levels. As expected, and because of the findings showing that STAT3 expression modifies AgRP expression [26], constitutive activation or deletion of STAT3 in this neuronal population regulates energy balance [14,55]. STAT3 deletion on AgRP neurons leads to a mild gain in body weight, and unlike the POMC neurons, this deregulation is due to effects on energy expenditure [55]. Overexpression of STAT3 in AgRP neurons is associated with increased locomotor activity, with no effects on food intake [14]. This is consistent with the recovery of locomotor activity after restoration of leptin levels in LepR knockout mice [14,56]. Furthermore, STAT3 has the effect of ameliorating impaired glucose homeostasis on mice exposed to a high-fat diet (HFD) [14], indicating the important role that STAT3 has in mediation of leptin effects. Altogether, these reports provide a clear implication that STAT3 is involved in the mechanism(s) by which leptin and insulin elicit responses by the POMC and AgRP neurons. Interestingly, overexpression of STAT3 in each separate neuron population results in different effects on energy balance and glucose homeostasis, indicating that synergy between the groups of neurons is necessary for a complete and appropriate response to leptin.

The PI3K/PDK1/FOXO1 signaling pathway acts to integrate leptin and insulin signals. The primary effector of this pathway, PI3K, has been shown to be a mediator of both hormones. PI3K is required for the actions of leptin and insulin to activate or inhibit POMC and AgRP neurons [18,20,24,37]. It has been demonstrated that the PI3Kp110β subunit has a fundamental role in regulating body weight, food intake, and glucose homeostasis through leptin and insulin signals [57]. Once more, attention has focused on the importance of PI3K in mediating the effects of both hormones; it became obvious that the mechanisms through which PI3K provokes changes in energy homeostasis were not clearly identified. Several studies report the main PI3K target in affecting energy homeostasis is PDK1. Recently, it was demonstrated that PDK1 is involved in regulating the ability of POMC and AgRP neurons to sense leptin and insulin inputs [17,58,59]. In POMC neurons, the deletion of PDK1 in mice promotes a gain in body weight associated with increased food intake [17,58]. As expected, mice with this ablation exhibit lower POMC expression [17,58], which is responsible for the changes in food intake seen in these animals. Furthermore, PDK1 deletion in POMC neurons affects glucose homeostasis. Administering leptin to these animals cannot reproduce the body weight loss provoked in control mice, suggesting that leptin action in POMC neurons is dependent upon PDK1 [58]. As this pathway plays an integrative role in leptin and insulin transduction signaling, it has been reported that PDK1 is involved in insulin action, too. The lack of PDK1 in POMC neurons leads to increased insulin sensitivity in adult mice and higher blood glucose levels in young and obese mice [17]. The deletion of PDK1 in this neuronal population has no effects on energy expenditure. On the other hand, in AgRP neurons, PDK1 was shown to regulate energy balance by provoking changes in food intake and energy expenditure [59].

As anticipated, and contrary to POMC effects, the deletion of PDK1 in the AgRP population promoted body weight loss as a consequence of diminished food intake and enhanced energy expenditure—specifically increased locomotor activity [59]. In addition, these mice were more sensitive to leptin. Chronic peripheral administration of leptin induced greater reductions in body weight and food intake [59]. PDK1 acts by activating AKT and by regulating many kinases and transcription factors. One of the most studied of these is FOXO1, which has been involved in hypothalamic neuropeptide expression and food intake. As stated above, when FOXO1 is activated by leptin in POMC neurons, it promotes POMC transcription [26], but conversely in AgRP neurons, it inhibits neuropeptide expression [27]. As expected, previous data from the two PDK1 knockout (KO) models described earlier are in agreement with FOXO1's effects on feeding. PDK1 deletion in POMC neurons impairs FOXO1 phosphorylation, and thus, this transcription factor remains in the nucleus to inhibit POMC expression [17,58]. In AgRP neurons, PDK1 ablation leads to nuclear accumulation of FOXO1, thereby decreasing food intake [59]. This data indicates that FOXO1 could be the terminal effector regulating leptin signaling; in fact, in these studies stimulation of FOXO1 rescues the phenotype. Its activation in PDK1ΔPOMC mice abolished this change in body weight, POMC expression and glucose homeostasis [17,58]. In AgRP neurons, FOXO1 activation, which leads to its exclusion from the nucleus, rescuing the phenotype seen in the AGRP:PDK1KO mice [59]. These studies emphasize the importance of FOXO1 as a regulator of leptin and insulin actions.

Several studies have described the specific action of FOXO1 in the regulation of food intake, energy expenditure, and glucose homeostasis. FOXO1 is a transcription factor that plays a role in regulating glyconeogenesis and glycogen breakdown. Leptin and insulin require FOXO1 signaling in the Arc to provoke their anorexigenic effects, and that leads to increased leptin and insulin plasma levels. Given these clues, the role of FOXO1 in POMC and AgRP neurons was investigated [22,26,60,61]. In POMC neurons, the deletion of FOXO1 leads to body weight loss associated with decreased food intake [60]. This change in feeding behavior is not associated with changes in POMC expression, but the authors linked it with increased levels of α-melanocyte-stimulating hormone (α-MSH), a proteolytic product of POMC, and they found that FOXO1’s actions on energy balance are mediated by carboxy peptidase E (CPE), a prohormone processing convertase [60]. No changes in energy expenditure were found in this model. Furthermore, FOXO1 expression in POMC neurons plays an important role in leptin and insulin action [60,61]. The deletion of this transcription factor in this POMC neuron increased leptin sensitivity [60]. Recently, an elegant report investigated the role of FOXO1 in AgRP neurons [22], demonstrating that FOXO1 signaling is required in AgRP neurons in order to regulate energy homeostasis. The specific ablation of FOXO1 in this neuronal population was associated with decreased food intake.
and locomotor activity [22]. Interestingly, body weight was not affected. In addition, it has been shown that a lack of FOXP1 in AgRP neurons affects glucose homeostasis. This animal model presents altered HGP due to enhanced insulin sensitivity in the AgRP neurons [22]. In this work, G protein-coupled receptor 17 (Gpr17) was identified as the downstream target through which FOXP1 affects energy homeostasis in the AgRP neurons [22].

Summarizing, the individual contributions of the most important factors of leptin and insulin signaling transduction pathways have been described. STAT3 and FOXP1, the terminal targets of both pathways are shown to play key roles in energy balance and glucose homeostasis and can act independently within either the POMC or AgRP neurons in the Arc. In spite of this, any of these factors are able to regulate the total overall effect of leptin and insulin with regard to: food intake, energy expenditure, or glucose metabolism. This indicates that there must be a cooperative and synergistic interaction between pathways in order to elicit responses by the AgRP and POMC neurons to mediate anorexigenic effects.

On the other hand, mTOR is a downstream target of the PI3K pathway, and its activation is dependent upon nutrient availability [25,29]. In addition to its role to sense glucose and leucine levels, it has been shown that this pathway mediates the anorexigenic effects of leptin [25,62,63]. Pharmacologic studies have demonstrated that mTOR inhibition by rapamycin blunts leptin effects on food intake [25]. Furthermore, these results were corroborated by other reports using transgenic models. Mice with a selective lack of S6K were unable to respond to leptin treatment, showing that this pathway is involved in the mechanism which leads to leptin resistance [63]. The activation of the mTOR pathway in POMC neurons attenuates body weight loss and decreases food intake after leptin treatment [64]. This data supports the idea that the mTOR pathway is an important mediator of the onset of leptin resistance. As stated earlier, AMPK mediates leptin action, and is inversely regulated by mTOR. A recent report demonstrated that leptin action on the AMPK pathway is dependent upon mTOR [65]. Leptin-mediated activation of mTOR promotes inhibition of the AMPK pathway and thus regulates feeding behavior [65]. This novel mechanism accentuates the importance of the mTOR pathway in mediating the effects of leptin.

Hypoxia inducible factor (HIF), closely related to mTOR, is a heterodimer formed by α and β subunits. Meanwhile the β subunit is constitutively expressed, the α subunit is modulated by oxygen availability. Under hypoxic conditions, HIFα is stabilized, and under normoxia, it is degraded [66,67]. It has been demonstrated that this transcription factor is involved in the control of glucose metabolism [68]. The deletion of HIF in POMC neurons curtails the changes on POMC expression caused by glucose injection, demonstrating that HIF mediates glucose action on POMC neurons [68]. Leptin’s effects are independent of HIF. AMPK and mTOR are nutrient sensors [23]. The activation of HIF prevents the effect on food intake provoked by AICAR, an AMPK activator. In the same way, the orexigenic effects of rapamycin are blocked by HIF activation [68]. Furthermore, decreased food intake caused by the activation of mTOR in the mediobasal hypothalamus is reduced in mice lacking HIF in their POMC neurons [68]. This report shows a potent action of HIF in glucose sensing, and that this effect is mediated in POMC neurons through the involvement of AMPK and mTOR signaling [68]. However, the role of HIF is not only in glucose sensing. In the same report, the authors propose that HIF acts as a therapeutic target to oppose obesity. The animal model used in this study is more sensitive to HFD; however, the constitutive activation of HIF protected these mice against diet-induced obesity [68].

**Potential role of autophagy in leptin’s and insulin’s mechanisms of action**

Macroautophagy (referred to as autophagy) is a novel mechanism involved in the maintenance of cellular homeostasis and it may also be involved in the mechanism of action for leptin. Autophagy is a process whereby cytoplasmic organelles are sequestered and delivered into lysosomes for degradation [69]. Although the role of autophagy has been tested in other tissues [70,71], this mechanism has only recently been studied in the hypothalamus. In the last year, some reports have suggested that autophagy plays a role in the regulation of food intake [69,72,73]. Mice that lack Atg7 (autophagy-related gene 7) selectively in POMC or AgRP neurons show altered energy homeostasis. In AgRP neurons, inhibition of autophagy promotes body weight loss associated only with increases in locomotor activity. In turn, they have a reduced rebound effect on food intake after fasting, indicating perhaps that these mice develop enhanced leptin sensitivity [69]. The deletion of Atg7 in POMC neurons leads to obesity, and changes in food intake and energy expenditure [72], supporting the idea that autophagy regulates energy homeostasis. In agreement with the findings in AgRP neurons, the inhibition of autophagy in POMC neurons provokes potent effects in glucose homeostasis. These mice exhibit leptin resistance, and this is related to the inability of leptin to activate and phosphorylate STAT3 [73]. Furthermore, the lack of Atg7 in POMC neurons increased blood glucose levels despite elevated insulin concentrations, indicating that these mice develop insulin resistance [72]. This data clearly demonstrates that autophagy plays a role in glucose homeostasis. In spite of this, the mechanism through which autophagy affects energy balance is still unknown; it has been proposed that the mTOR pathway can act to mediate autophagy-dependent changes in energy homeostasis [69,72].

In conclusion, in recent years, the individual contribution of each significant component of the JAK/STAT3 and PI3K pathways in the regulation of energy balance and glucose homeostasis has been described. Furthermore, studies of STAT3 and FOXP1, the terminal effectors of these pathways, have shown that they play vital roles in leptin and insulin actions. The POMC and AgRP neurons of the Arc of the hypothalamus have been identified as key central targets of leptin and
insulin action. New targets, such as mTOR, have been detected to be potent mediators of leptin and insulin signaling, and new mechanisms, such as autophagy, have been discovered to be involved in energy and glucose homeostasis. Despite this, more efforts are needed to understand the individual contribution of each factor discussed here and the global mechanisms by which they operate with the ultimate goal of obtaining a comprehensive understanding of leptin and insulin resistance. This would thereby advance the development of future therapeutic treatments to combat obesity and its associated metabolic syndrome.

References

4 Mohr B: Hypertrophie der Hypophyse und Druck auf die Hirnwindungen. Wisschr ges Heilk 1840;6:565–571. [German]
6 Aschner B: Über die Funktion der Hypophyse. Arch f. d. ges Physiol 1912;146:1–146. [German]
The role of the hypothalamus in the maintenance of energy balance and peripheral glucose control


SECTION VII

Clinical trial evidence in diabetes
Key points

- Type 1 diabetes (T1DM) is a progressive disease with genetic predisposition.
- T1DM is characterized by immunologically mediated β-cell destruction.
- Immune intervention studies have been conducted in recent-onset T1DM in an effort to interrupt the disease process and hinder the anticipated progressive decline in β-cell function.
- Some interventions have shown at least transient effect in recent-onset T1DM.
- Immune intervention studies have also been conducted in individuals at risk of T1DM in an effort to interrupt the disease process and prevent progression to clinical T1DM.
- No prevention study, as yet, has been unambiguously successful.

Introduction

Type 1 diabetes (T1DM) occurs when there is sufficient reduction of β-cell mass and function to require insulin therapy to control glycemia. Since the early 1970s, when islet cell antibodies (ICA) were first noted [1], and an association with human leukocyte antigen (HLA) was first described [2], T1DM has been characterized as a distinct disease from type 2 diabetes (T2DM). In the 1980s, Eisenbarth characterized T1DM as a slowly progressive disease that had a genetic predisposition and was characterized by immunologically mediated β-cell destruction [3]. This was certainly consistent with spontaneous diabetes in two animal models, the BioBreeding (BB) rat [4] and the nonobese diabetic (NOD) mouse [5]. Over the years, multiple studies in these rodent models have shown that diabetes can be prevented or stabilized by a vast variety of immunologic approaches [6]. This has led to multiple studies in human beings using immune interventions designed to prevent T1DM or to slow the decline in β-cell function characteristic of the disease process. This chapter summarizes the human studies that have been conducted, focusing in particular on randomized controlled clinical trials.

Most studies have been conducted in people with recently diagnosed T1DM, in an effort to preserve β-cell function. The earliest studies used clinical “remission” (variously defined) as the primary outcome measure. Most studies in the past two decades, however, have used as the primary outcome measure stimulated C-peptide as an index of β-cell function, usually in response to a standardized meal (“mixed meal tolerance test” (MMTT)), although some have used C-peptide response to glucagon administration (“glucagon stimulation test” (GST)). One study ambitiously used as its primary outcome measure C-peptide response during a hyperglycemic clamp followed by a GST.

Studies in established T1DM

Although prevention of T1DM is the logical ultimate goal of immune intervention studies, most studies to date have been conducted in recent-onset T1DM in an attempt to impede the disease process and preserve β-cell function. The first such attempt at immune intervention in T1DM was a study conducted in the late 1970s [7]. Multiple groups then tested a variety of approaches, most of which were not randomized controlled trials, and even those that were randomized were so small that there were insufficient numbers of subjects for valid conclusions [8,9]. Many early studies used either no control subjects, historical comparison groups, or concurrent nonrandomized comparison groups.

A pilot study, led by Calvin Stiller, and published in Science in 1984 [10], stimulated the development of multiple studies evaluating the effects of the immunosuppressive drug cyclosporin in T1DM [11–13]. Cyclosporin is a calcineurin inhibitor that interferes with the activity and growth of T-lymphocytes. Two large randomized controlled clinical trials—one from the French Cyclosporin Study Group [11] and the other from the
During the 1980s and 1990s, several other interventions were evaluated. Three studies were conducted using azathioprine, an immunomodulatory drug inhibiting purine synthesis [16–18]. One of these used azathioprine together with glucocorticoids [16]. None of the three studies showed convincing results that might have stimulated the field. Another study was conducted with linomide, an immunomodulatory drug inhibiting pyrimidine synthesis [19]. In this randomized controlled trial, there was improvement in the GST in those subjects who had detectable residual β-cell function at entry. Two randomized controlled trials were conducted by vaccination with BCG (bacillus Calmette-Guérin strain of Mycobacterium bovis) [20,21]. These failed to demonstrate a beneficial effect on stimulated C-peptide, and in the larger study, there was a trend toward a more rapid decrease in C-peptide levels in BCG-treated subjects [21]. Oral insulin was also used in three studies in recent-onset T1DM, all essentially without beneficial effect [22–24]. Two small pilot studies reported in the early 1990s were the first to use monoclonal antibodies in an attempt to alter the course of T1DM, one using an anti-CD5-immunotoxin [25] and the other using an anti-CD4 [26]. These heralded the testing of two anti-CD3 monoclonal antibodies, which have been used in multiple controlled clinical trials.

Several studies have been conducted with anti-CD3 monoclonal antibodies, which target T-lymphocytes, particularly CD4+T-lymphocytes, but with relative sparing of T-regulatory cells. Two anti-CD3 antibodies have been used—teplotizumab and otelixizumab [27–35]. The course of treatment was just 14 days for teplotizumab and 6 days for otelixizumab. The phase 2 studies with both drugs showed beneficial effect on C-peptide [27–29], with effects sustained for 2 years after the initial treatment [28]. With otelixizumab, beneficial effects on A1c and insulin dose were seen as long as 4 years after the initial 6-day treatment course, although C-peptide response was not measured at that time [30]. Unfortunately, the phase 3 studies did not meet their primary outcome criteria. For teplotizumab, the primary outcome was the combination of HbA1c <6.5% and insulin dose <0.5 units kg⁻¹ day⁻¹, a dichotomous measure which was arbitrarily chosen [31]. Yet, when C-peptide was examined there was evidence of efficacy both at 1 year [31] and at 2 years [32] following 14-day courses of teplotizumab at entry and at 26 weeks into the study. For otelixizumab, the phase 3 study used a dose that was one-sixteenth (total of 3.1 mg over 8 days) of that used in the phase 2 study (total of 48 mg), in an effort to avoid any side effects [33]. Not only were side effects obviated, but so was efficacy. As a consequence, a formal dose-ranging study (NCT02000817) is currently underway in Belgium, using doses of 9, 18, 27, and 36 mg over 6 days.

There have been two other anti-CD3 studies with teplotizumab. In one, the DELAY trial, rather than enrolling subjects with recent-onset T1DM, subjects with T1DM for 4–12 months after diagnosis were enrolled [34]. In this study, there was a beneficial effect on C-peptide secretion 12 months after enrollment, accompanied by lower insulin doses. The other study with teplotizumab, the AbATE Study, describes a group of “responders” to treatment, defined as those who maintained C-peptide better than the randomized but untreated comparison group at 24 months [35]. Interestingly, there was little or no decline in C-peptide in the responder group, which was about 45% of subjects treated with anti-CD3. This is particularly interesting because in the other anti-CD3 trials, the C-peptide response was transient, with a decline after 6–12 months in parallel to the comparison group. One interpretation could be that the treated subjects were a mixed group, with some being responders and others nonresponders, thus confounding interpretation.

There have been a number of other immune intervention studies conducted in the past 10 years, many by two NIH networks, Type 1 Diabetes TrialNet and the Immune Tolerance Network (ITN). Two of the TrialNet studies showed beneficial results in terms of C-peptide response on a MMTT. One of these used the anti-CD20 monoclonal antibody rituximab, which targets B-lymphocytes [36]. Treatment was weekly for 4 weeks, with the primary outcome measured at 1 year [36] and follow-up at 2 years [37]. Although there was increased C-peptide, the effect appeared to be transient with calculated delay in decline, in comparison to the placebo group, being 8.2 months [37].

One other TrialNet study also showed a beneficial effect using the co-stimulation blocker abatacept [38]. Treatment was monthly for 2 years, with the primary outcome measured at 2 years [38] and follow-up at 3 years, 1 year after cessation of treatment [39]. Although there was a difference in C-peptide response to a MMTT favoring abatacept, there was a decline from the inception, parallel to the placebo group [38]. The difference between the groups suggested that there was a delay of decline of 9.5 months [39].

Another study conducted by TrialNet evaluated mycophenolate mofetil (MMF) (a drug that inhibits purine synthesis in the proliferation of both B- and T-lymphocytes) with or without initial treatment with the anti-CD25 monoclonal antibody daclizumab [40]. This study was halted prior to completion because the Data Safety and Monitoring Board (DSMB) determined that the C-peptide results were no different between each intervention and its control group and that it was futile to expect a difference to emerge.
TrialNet also studied the anti-IL-1β monoclonal antibody canakinumab [41]. Interleukin-1β (IL-1β) is a pro-inflammatory cytokine that recruits effector T-lymphocytes in inflamed tissues, and which also has direct toxic effects on β cells. Unfortunately, the use of canakinumab by itself, given monthly for 1 year, failed to show any difference in MMTT-stimulated C-peptide [41]. Because a study of anakinra, an IL-1 receptor antagonist, was concomitantly completed by the European AIDA Study Group, and also failed to show benefit, both studies were included in a single paper [41]. In contrast, a small pilot study conducted in Buffalo, NY using etanercept, a blocker of tumor necrosis factor (TNF), another pro-inflammatory cytokine, suggested that C-peptide was better in treated than control subjects [42]. It may be that anti-inflammatory agents such as anti-IL-1β or anti-TNF, targeting innate immunity, are best used in combination with agents targeting adaptive immunity, such as anti-CD3, anti-CD20, or co-stimulation blockade.

A recent pilot phase 1 safety study used alpha-1-antitrypsin (AAT) [43]. No safety issues were identified. The study did show that AAT was associated with a down-modulation of IL-1β, which the authors suggested may be proof of benefit for T1DM [43].

ITN conducted a trial using thymoglobulin, a polyclonal anti-lymphocyte preparation that broadly suppresses immune responses [44]. Unfortunately, the results failed to show beneficial effect on C-peptide response.

Another ITN trial evaluated alefacept, a dimeric fusion protein targeting CD2, which is expressed on most T-lymphocytes, but with highest expression on effector memory T-lymphocytes cells and central memory T-lymphocytes, and most prominently on highly pathogenic armed effector T-lymphocytes [45]. Subjects received 12 weekly injections, followed by a 12-week pause, and then another 12 weeks of treatment. Unfortunately during the study, the manufacturer ceased production of the drug and withdrew it from the market, resulting in a smaller than planned sample size. In the treated group, there was no decline in C-peptide from baseline. Yet, the primary outcome, C-peptide during the first 2 hours of a MMTT, failed to meet statistical significance, although a secondary outcome, C-peptide during the full 4 hours of a MMTT, did show benefit versus placebo. Moreover, there was a clear decrease in central memory T-lymphocytes whilst regulatory T-lymphocytes were unchanged [45]. The beneficial effect on C-peptide (using the 4-hour MMTT), together with the pattern of alteration of T-lymphocytes, suggests that targeting central memory T-lymphocytes may be a useful mechanism worthy of further study in T1DM.

Antigen-based therapies are considered highly desirable because they are unlikely to have off-target adverse effects. As noted earlier, oral insulin was used in three studies in recent-onset T1DM without benefit [22–24]. Another diabetes antigen is glutamic acid decarboxylase (GAD). A GAD vaccine, with GAD coupled to the adjuvant aluminum hydroxide (GAD-Alum) has been tested in a number of clinical trials. Although an initial report of a phase 2 study suggested benefit, at least in those subjects enrolled early after diagnosis [46], this was not confirmed in a TrialNet study [47] nor in a phase 3 study conducted by the manufacturer [48].

Other antigen-based studies, using an insulin B-chain altered peptide ligand [49], or a plasmid-encoded proinsulin [50], also failed to show benefit. Pilot safety studies with a proinsulin cytokine showed no harm [51], and additional safety studies are underway (NCT01536431) with both a single peptide, and plans to study a combination of proinsulin peptides. Another pilot study is using alpha methyldopa, which has been shown to prevent antigen presentation by HLA-DQ8 (NCT01883804). A 24-amino acid peptide derived from heat shock protein 60, called DiaPep277, has been studied in multiple phase 2 trials [52–56] and in a recent phase 3 trial [57,58]. Although there appeared to be promising results from the first phase 2 trial [52,53], results from the other phase 2 trials [54–56] were conflicting. The phase 3 trial gave mixed results, with improved C-peptide versus placebo when stimulated with a GST, but no difference between groups with the MMTT [57,58]. This leaves the results ambiguous.

ITN and TrialNet conducted an open label phase 1 study using a combination of interleukin-2 (IL-2) and rapamycin [59]. The study was halted due to an apparent acute decline in C-peptide in the subjects. It is unclear whether this was due to the high dose of IL-2 used or to adverse effects from the rapamycin. More recently, a safety study has been reported with low-dose IL-2, with no safety issues emerging, including no decline in C-peptide [60]. Further studies with low-dose IL-2 are underway (NCT01862120).

A small pilot study has been done with the combination of low-dose antithymocyte globulin (ATG) and granulocyte colony stimulating factor (GCSF) in patients with diabetes of 4 to 24 months’ duration [61]. It demonstrated preservation of β-cell function. As a consequence, a fully powered study with this combination is being initiated in recently diagnosed T1DM.

Another recent combination study evaluated sitagliptin and lansoprazole in patients with recent-onset T1DM [62]. The rationale of this study was that a DPP-4 inhibitor (sitagliptin) would increase serum levels of GLP-1, while a proton pump inhibitor (lansoprazole) would increase serum levels of gastrin. In experimental animals, GLP-1 and gastrin have been shown to increase β-cell mass and function. The hope was that this combination would be beneficial in human T1DM. Unfortunately, there was no difference in β-cell function when comparing the treated subjects and control subjects [62].

A variety of cellular-based approaches have also been attempted. Autologous umbilical cord blood infusion, either alone or followed by docosahexaenoic acid and vitamin D supplementation, has been studied in several small pilot studies [63–66]. This therapeutic approach has been found to be safe. Unfortunately, there has been no beneficial effect on β-cell function.

Autologous in vitro manipulated potentially tolerogenic dendritic cells have been studied and have been found to be safe [67]. Some of these dendritic cells were antisense DNA-treated
co-stimulation-impaired immunoregulatory dendritic cells, while some were unmanipulated. However, in spite of the fact that this was a small pilot study, there was the suggestion of upregulation of the frequency of a potentially beneficial B-lymphocyte population [67]. It is interesting that another group studied dendritic cells in vitro and found that they had the capability to show tolerance induction to insulin or GAD65, but have not yet treated individuals with these cells [68].

Regulatory T-lymphocytes (T-regs) have also been explored as potential therapy for T1DM [69–71]. One group has studied these in children with T1DM and has shown that repetitive administration of T-regs is safe and may even prolong survival of β cells in T1DM [69,70]. Another group conducted a phase 1 dose-escalation safety study in 14 subjects with T1DM, using autologous polyclonal T-regs that had been expanded from peripheral blood [71]. Infusions were well tolerated with no safety concerns, and infused T-regs were detectable for over 6 months. Subjects exhibited stable C-peptide for as long as 2 years [71]. As a consequence, a full phase 2 study to evaluate safety and efficacy is being initiated.

One therapeutic strategy that has been hailed as having achieved insulin independence is the combination of high-dose immunotherapy, generally with cyclophosphamide and anti-thymocyte globulin, together with nonmyeloablative autologous hematopoietic stem cell therapy (AHSCT) using CD34+ cells isolated from bone marrow [72–77]. The first study, conducted in Brazil, was initially reported in 2007 [72], with updated information reported in 2009 [73,74]. Studies from Poland [75] and China [76] were subsequently reported, and there has been a recent summary of the Polish and Chinese data, including an additional Chinese site [77]. The remarkable finding was that a substantial number of subjects achieved insulin independence, some lasting for as long as 4 years [73,77]. The dilemma is that there were substantial side effects, including a death from Pseudomonas sepsis [77], and that the death rate in other disease states with AHSCT can be as high as 10% [78]. Moreover, the characterization of some of the subjects was such that it was not clear that all had autoimmune T1DM. An important potential issue is whether the interpretation of the mechanism responsible for success is that (a) the effectiveness of AHSCT relies on immunoregulatory properties of CD34+ cells, or that (b) the high-dose immunosuppression is responsible for the beneficial effects with the AHSCT serving to rescue the subject from aplastic anemia and bone marrow failure. If the former, better ways of isolating and using CD34+ cells may be required, while if the latter, safer ways of using immunotherapy would be required.

From the hints of success, it is apparent that some form of combination therapy will be necessary to alter the course of recently diagnosed T1DM [79–81]. It is the author’s bias that this will require different modalities, including modulation of the adaptive immune system, suppression of innate immunity, enhancement of protective immunity, all of this possibly in the context of antigen specificity, and some modality that enhances β-cell function and/or enhances β-cell survival [79].

### Studies designed to prevent T1DM

Ideally, it would be desirable to prevent T1DM, since earlier therapy may better alter the immune processes, and with earlier therapy there is a greater likelihood that β-cell mass has not been severely compromised. On the other hand, given that clinical disease has not yet emerged, therapies tested must be safe and not likely to have adverse effects. There are two categories of prevention trials—primary prevention and secondary prevention [82]. Primary prevention trials perform genetic screening at birth, and conduct trials in infants with genetic risk of T1DM. Secondary prevention trials involve strategies that either follow a birth cohort with genetic risk until the development of signs of autoimmunity, or screen for autoimmunity amongst relatives (usually first-degree relatives) of individuals with T1DM. Those with genetic risk who are followed from birth for autoimmunity have an extraordinarily high probability of developing T1DM once they develop two or more diabetes autoantibodies [83].

All of the primary prevention trials conducted to date have used dietary interventions. Following epidemiologic studies and meta-analyses that suggested that early introduction of cow’s milk may serve as a trigger of T1DM [84,85], a pilot study was initiated to determine if it was feasible to conduct a study in which cow’s milk-based formula would not be used [86]. The trial studied babies with high genetic risk, and encouraged breast feeding, but at the time of weaning, randomized infants to receive either a casein hydrolysate formula or a conventional cow’s-milk formula (control) whenever breast milk was not available during the first 6–8 months of life. Initial reports were encouraging and eventually even suggested that there might be a reduction of autoimmunity [86]. The feasibility of conducting such trials led to the rather ambitious “Trial to Reduce Incidence of Diabetes in the Genetically at Risk” (TRIGR), which registered over 5000 newborns in 15 countries and randomized 2160 newborns to receive either casein hydrolysate formula or cow’s-milk formula [87,88]. Unfortunately, TRIGR did not confirm that there was a reduction in autoimmunity with the casein hydrolysate formula [89].

Other dietary interventions have also been studied. A pilot study, the Finnish Dietary Intervention Trial for the Prevention of Type 1 Diabetes (FINDIA), sought to establish whether an infant formula free of bovine insulin may reduce T1DM autoimmunity [90]. The FINDIA formula resulted in a reduced risk of appearance of one or more antibodies when compared with cow’s-milk formula [90].

Observational studies suggested that age of introduction of solid food, particularly gluten-containing foods or cereals, affects the development of autoimmunity in children with genetic risk of T1DM [91,92]. In a pilot study
in antibody-positive children, β-cell function seemed to be improved by deprivation of gluten for 6 months [93]. This led to the BABYDIET study, a randomized controlled trial, which evaluated whether delayed exposure to gluten reduces the risk of diabetes autoimmunity [94]. Unfortunately, delaying gluten exposure until 12 months of age, although safe, did not reduce the risk for diabetes autoimmunity in children at genetic risk of T1DM [94].

Observational studies also suggested that dietary intake of omega-3 fatty acids is associated with reduced risk of islet autoimmunity in children at increased genetic risk for T1DM [95]. The TrialNet Nutritional Intervention to Prevent (NIP) Type 1 Diabetes Pilot Trial assessed the feasibility of nutritional supplementation with the omega-3 fatty acid docosahexaenoic acid (DHA) during the last trimester of pregnancy and the first few years of life [96]. During the relatively brief follow-up, there was no change in autoimmunity although other changes in anti-inflammatory factors suggested that omega-3 fatty acids may have some effects that would need to be further evaluated [97].

The other potential dietary intervention that has received much discussion is vitamin D supplementation. Vitamin D prevents both insulinitis and T1DM in mouse models of T1DM [98,99]. Retrospective studies have noted apparent beneficial effects of vitamin D supplementation in early life on subsequent lifetime risk of T1DM [98]. Other studies have not found that to be the case [100], although a meta-analysis of four case-control studies and one cohort study did suggest reduced risk with early vitamin D supplementation [101]. A feasibility study has shown that it is possible to recruit babies from the general population for identification of genetic risk and enrollment to a randomized controlled prevention trial of vitamin D supplementation [102]. Such a full-scale trial is clearly needed to clarify whether or not vitamin D supplementation is truly beneficial.

Finally, another primary prevention strategy under preliminary evaluation is the mucosal presentation of antigen, either orally or nasally. The PrePoint study is evaluating doses of antigen in very high-risk children with genetic markers, to determine whether this may prevent the emergence of islet autoantibodies [103].

It may be argued that one should test vaccines for primary prevention—either vaccines against putative viral or other infectious triggers, or antigen-specific vaccines. To date, no such studies have been initiated.

Several secondary prevention trials have been conducted in subjects who already have diabetes-related autoantibodies. Thus, the outcome measure in these trials is the development of T1DM. Again, like in primary prevention trials, because the research subjects do not have a clinical diagnosis, any intervention to be tested should be safe. Nicotinamide, a water-soluble vitamin (B6) derived from nicotinic acid, was studied in two large multicenter randomized controlled clinical trials in relatives of individuals with T1DM found to be at risk of the disease because of presence of diabetes autoantibodies [104,105]. These two studies—the German (Deutsch) Nicotinamide Diabetes Intervention Study (DENIS) and European Nicotinamide Diabetes Intervention Trial (ENDIT)—did not prevent or delay development of T1DM [104,105].

Antigen-based strategies have also been studied for secondary prevention of T1DM. The Diabetes Prevention Trial-Type 1 (DPT-1) Study Group conducted two studies evaluating insulin as a potential antigen-based therapy, enrolling relatives of individuals with T1DM screening positive for islet cell antibodies. The two studies used: (i) parenteral (or injected) insulin in individuals with a projected 5-year risk of at least 50% [106], and (ii) oral insulin in individuals with a projected 5-year risk of 25–50% [107]. The parenteral insulin trial tested not only whether insulin as an antigen could modulate immunity, but also whether insulin use might result in “β-cell rest”, thus decreasing presentation of antigen to the immune system. However, the rate of development of diabetes was essentially the same in both the treated and control arms of the DPT-1 parenteral insulin trial [106].

In the DPT-1 oral insulin trial, the rate of development of diabetes was essentially the same in both the treated and placebo arms [107]. However, in a post hoc analysis, a subgroup of subjects—those with higher titers of insulin autoantibodies at the time of enrollment—appeared to show a beneficial effect of oral insulin with a projected delay of 4.5 to 5 years in onset of T1DM [107]. Long-term follow-up of the DPT-1 oral insulin cohort showed that even after administration of oral insulin was ceased, effects were maintained [108]. As a consequence of these observations, another oral insulin trial is being conducted by Type 1 Diabetes TrialNet, which is enrolling subjects with characteristics similar to the subgroup that showed delay in T1DM in the DPT-1 trial.

A Belgian trial also evaluated parenteral insulin, and also failed to show a delay in development of T1DM [109].

A trial in Finland, the Type 1 Diabetes Prediction and Prevention Study (DIPP) was conducted in newborns with genetic risk of T1DM (but without a known relative with the disease), followed prospectively for development of diabetes autoantibodies, after which they were treated with nasal insulin or placebo [110]. Unfortunately, again there was no difference in rate of progression to T1DM [110].

Nasal insulin was also studied for safety in the Intranasal Insulin Trial (INIT 1) [111]. No adverse effects were seen. A subsequent study of nasal insulin in patients who already had T1DM, demonstrated that the interferon-γ response of T-lymphocytes to proinsulin was suppressed after nasal insulin [112]. As a consequence, another nasal insulin trial is being conducted by the Australian Diabetes Vaccine Development Centre (DVDC).

In a small pilot trial using low-dose cyclosporine (six subjects compared to nine historical controls) in relatives of T1DM patients with islet cell antibodies and some metabolic
impairment, two of the cyclosporine-treated subjects remained without T1DM for 47 and 57 months [113]. This was not further pursued due to potential cyclosporine toxicity, as discussed earlier.

Several other prevention studies are currently underway. These include a study of GAD vaccination, the DIAPREV-IT (Diabetes Prevention - Immune Tolerance) Study in Southern Sweden (NCT01122446), a TrialNet study with the anti-CD3 monoclonal antibody teplizumab (NCT01030861), and another TrialNet study with the co-stimulation blocker abatacept (NCT01773707).

**Conclusions**

Hundreds of interventions have been shown to prevent or reverse immune-mediated diabetes in rodent models. Yet, none has been unambiguously successful in human beings, either to prevent T1DM, or to stabilize β-cell function over a protracted timeframe. Nonetheless, several human studies have had transient beneficial effects without major safety issues. This raises the prospect that combination therapies may yet prove beneficial [79–81].

**References**


87 TRIGR Study Group: Study design of the Trial to Reduce IDDM in the Genetically at Risk (TRIGR). *Pediatric Diabetes* 2007;8:117–137.
Prevention of type 2 diabetes

Thomas Yates and Melanie J. Davies
Diabetes Research Centre, College of Medicine, Biological Sciences and Psychology, University of Leicester, Leicester, UK; National Institute for Health Research Leicester-Loughborough Diet, Lifestyle, and Physical Activity Biomedical Research Unit, UK

Key points
- Diabetes prevention trials and initiatives have been focused on those with prediabetes, defined by the presence of impaired glucose tolerance (IGT) and/or impaired fasting glucose (IFG).
- If left untreated, up to 70% of those with prediabetes will develop type 2 diabetes (T2DM) over the course of their lifetime.
- Unhealthy lifestyle practices attributable to modern industrialized environments have been shown to account for 80–90% of all cases of T2DM.
- Lifestyle interventions studies have been shown to reduce the risk of developing T2DM by 30–60% in those with prediabetes.
- Some pharmaceutical agents, particularly metformin, have also been shown to be effective in the prevention of T2DM.
- Lifestyle interventions and metformin are likely to be cost-effective in those with prediabetes or even incur a modest cost saving in some subgroups.
- A two-stepped approach, whereby risk scores are used to identify moderate- to high-risk individuals and blood tests are then used to confirm risk (prediabetes) status, is gaining consensus as the cost-effective method of identifying individuals who are suitable for a diabetes prevention program in routine care.
- Diabetes prevention programs focused on lifestyle change have been successfully translated into routine clinical care across diverse healthcare systems internationally.

Prediabetes

Unlike many other chronic diseases, T2DM is not heralded by a discrete physiologic event which defines diagnoses; rather it sits at one end of a continuous glucose control spectrum with normal glucose control at the other. In-between these two boundaries exists a region of abnormal glucose control which is already characterized by concomitant insulin resistance and β-cell dysfunction but does not yet reach the criteria for T2DM [1]. Those diagnosed with T2DM typically spend an extended period in this region of impaired glucose regulation, sometimes for more than a decade, before progressing to T2DM [1]. This affords an identifiable window of opportunity with which to apply primary prevention initiatives aimed at reducing insulin resistance and/or improving β-cell function to slow or reverse progression to T2DM.

Higher than normal glucose levels that are below the threshold for T2DM are defined in various ways depending on the glucose measurement selected and the criteria applied (see Table 38.1). Various terminologies have been used to indicate the presence of at least one glucose measure in this higher than normal range, including prediabetes (the most widely used), nondiabetic hyperglycemia and impaired glucose regulation. Although we recognize the controversy surrounding the term prediabetes, and acknowledge the fact that not all individuals with prediabetes will develop T2DM, where expedient we use this term for brevity and comprehensibility.

Impaired glucose tolerance and impaired fasting glucose

Prediabetes has traditionally been defined by the presence of impaired glucose tolerance (IGT) and/or impaired fasting glucose (IFG). IGT is characterized by elevated postprandial glucose and is identified by means of an oral glucose tolerance test (OGTT) whereby plasma glucose levels are measured 2 hours after a standardized glucose load of 75g taken in the fasting state. Both the World Health Organization (WHO) and the American Diabetes Association (ADA) recognize 2-hour post challenge glucose (henceforth referred to as 2-h glucose) levels of greater than or equal to 7.8 mmol L\(^{-1}\) and less than 11.1 mmol L\(^{-1}\) as indicating IGT (see Table 38.1) [2,3].

IFG is characterized by elevated fasting glucose levels. Compared to IGT, there is less international unanimity over the classification of IFG, with ADA recommending values of greater than 5.5 mmol L\(^{-1}\) and less than 7.0 mmol L\(^{-1}\), and WHO...


Table 38.1  Current WHO and ADA criteria for impaired fasting glucose, impaired glucose tolerance and HbA1c defined high risk states

<table>
<thead>
<tr>
<th>Current WHO criteria</th>
<th>Impaired glucose tolerance</th>
<th>Fasting glucose</th>
<th>2-h post challenge glucose</th>
<th>HbA1c</th>
</tr>
</thead>
<tbody>
<tr>
<td>[2,17]</td>
<td>&lt;7 mmol L⁻¹</td>
<td>≥7.8 mmol L⁻¹ and &lt;11.1 mmol L⁻¹</td>
<td>&lt;6.5% (if measured)</td>
<td></td>
</tr>
<tr>
<td>Impaired fasting glucose</td>
<td>≥6.1 mmol L⁻¹ and &lt;7.0 mmol L⁻¹</td>
<td>&lt;7.8 mmol L⁻¹</td>
<td>&lt;6.5% (if measured)</td>
<td></td>
</tr>
<tr>
<td>HbA1c defined high state</td>
<td>No recommendation given</td>
<td>No recommendation given</td>
<td>No recommendation given</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Current ADA criteria</th>
<th>Impaired glucose tolerance</th>
<th>Fasting glucose</th>
<th>2-h post challenge glucose</th>
<th>HbA1c</th>
</tr>
</thead>
<tbody>
<tr>
<td>[3]</td>
<td>&lt;7 mmol L⁻¹</td>
<td>≥7.8 mmol L⁻¹ and &lt;11.1 mmol L⁻¹</td>
<td>&lt;6.5%</td>
<td></td>
</tr>
<tr>
<td>Impaired fasting glucose</td>
<td>≥5.6 mmol L⁻¹ and &lt;7.0 mmol L⁻¹</td>
<td>&lt;11.1 mmol L⁻¹</td>
<td>&lt;6.5%</td>
<td></td>
</tr>
<tr>
<td>HbA1c defined high state</td>
<td>&lt;7 mmol L⁻¹</td>
<td>&lt;7.8 mmol L⁻¹</td>
<td>≥5.7 % and &lt;6.5%</td>
<td></td>
</tr>
</tbody>
</table>

Recommendation values of greater than or equal to 6.0 mmol L⁻¹ and less than 7.0 mmol L⁻¹ (see Table 38.1) [2,3]. IFG and IGT represent distinctive phenotypes with differing pathophysiologies. IFG is predominantly characterized by hepatic insulin resistance, near normal peripheral insulin sensitivity and severely impaired early-phase insulin response following a glucose load [1,4]. IGT is characterized by peripheral insulin resistance, near normal hepatic insulin sensitivity and impaired early-phase and late-phase insulin response following a glucose load [1,4]. Both IGT and IFG are associated with some degree of β-cell dysfunction. Indeed, studies have reported severe impairments in β-cell function in those with IGT and/or IFG with up to 80% decreased function with indirect measures and 50% decreased β-cell volume with autopsy findings [1,5–7].

In most countries around 15% of adults have IGT or IFG based on WHO criteria [2,8]; this figure rises in some minority populations and with age (in some elderly populations up to 1 in 2 individuals are estimated to have prediabetes [9]). Of those with IGT or IFG, an estimated 4–10% develop T2DM per year, with the highest rates (up to 20%) seen in those with both IGT and IFG [1,2,8]. Evidence from prospective studies suggests that around 25–40% of individuals with prediabetes go on to develop diabetes over a 3–8-year period and, according to an ADA expert panel, up to 70% will eventually develop T2DM over the course of their lifetime [10–13]. Rates closer to 90% have been observed in a long-term (20-year) follow-up of a Chinese population with IGT [14].

The risk of cardiovascular disease and premature mortality is also significantly elevated in individuals with IGT and IFG [2,8]. For example, in the DECODE trial, people with IGT were 50% more likely to die of cardiovascular disease than people with normal blood glucose control [15]. Furthermore, as there were four times more people with IGT as with T2DM, there were more premature deaths attributable to IGT than to diabetes [15,16].

HbA1c and prediabetes

Since WHO revised the criteria for the diagnosis of T2DM to include HbA1c [17], there has been continued discussion around the inclusion of HbA1c in the definition of a high-risk prediabetes category analogous to that of IGT or IFG. This is highly clinically relevant as it is potentially burdensome and confusing to define categories on a continuous glucose control spectrum with different measures. Although WHO found insufficient evidence for the use of HbA1c in the definition of prediabetes, recent statements from the American Diabetes Association and an international expert committee recommended that HbA1c can be used to signify a high-risk state at levels of 5.7–6.4 % and 6.0–6.4%, respectively [3,18]. NICE have also recently recommended that HbA1c levels of 6.0–6.4% can be used as an alternative to fasting or 2-h glucose in the identification of prediabetes. Although clinically useful, the effect that this change in classification may have on risk status is unclear; evidence from the UK and elsewhere suggests there is significant discordance in individuals identified with prediabetes through HbA1c or traditional criteria [19,20].

Lifestyle interventions

Unhealthy lifestyles and T2DM are inextricably linked, with the former being the primary causal agent of the latter. Unhealthy lifestyle practices attributable to modern industrialized environments have been shown to account for 80–90% of all cases of T2DM [21]. It has been speculated, through the so-called “thrifty genotype” and “thrifty phenotype” hypotheses [22,23], that phenotypes susceptible to obesity and insulin resistance were bestowed with survival advantages throughout the vast majority of human evolution, which was characterized by the harsh environments associated with hunter-gatherers and early subsistence agriculturists. However, in modern industrialized environments where the traditional cyclical nature of food availability has flat-lined at the highest possible level and the need for purposeful physical activity has been all but engineered out of our daily lives, these previously advantageous phenotypes have become highly deleterious. Given these considerations, lifestyle interventions have traditionally formed the primary therapeutic target in diabetes prevention programs.
Randomized controlled trials have confirmed the primacy of lifestyle in the prevention of T2DM, with lifestyle interventions reducing the risk of progressing to T2DM by 30–60% in those with prediabetes [24,25]. For example, the Finnish Diabetes Prevention Study (DPS) found that the risk of T2DM was reduced by 58% in those given lifestyle counseling compared to control conditions over a 3-year period [26]. Similar findings were seen in the USA in the Diabetes Prevention Study (DPP) [27]. These results have also been replicated across diverse countries and cultures including India [28], Japan [29], and China [30]. Table 38.2 details the main characteristics of these trials.

Importantly, the benefits of successful lifestyle change programs have also been shown to be lasting over the longer term. DPS, DPP, and the China Da Qing prevention study have all found sustained reductions in the incidence of T2DM relative to the control group over 7–20 years of follow-up [14,31,32]. These results were achieved even though the active lifestyle intervention discontinued after the initial study period. Therefore it would appear that once individuals are enabled to successfully change and self-regulate their health behavior, benefits can be accrued long after behavioral counseling has ceased.

Lifestyle interventions aimed at the prevention of T2DM in those with prediabetes have all been based on the promotion of moderate to vigorous intensity physical activity, a healthy diet, and weight loss for overweight or obese individuals or weight maintenance for normal-weight individuals (see Table 38.2). Each of these interventional components has independent effects on the risk of T2DM.

**Obesity**

Weight loss, of between 5 to 10% of initial body weight and maintenance, was a core component of all successful diabetes prevention programs. Although obesity and T2DM are symbiotically linked, the pathophysiology of how adiposity causes the development of T2DM is not fully understood; however, several mechanisms have been postulated. Adipose tissue, especially the tissue surrounding internal organs, is known to be an active endocrine organ that secretes a variety of pro-inflammatory adipokines, which act at both the local and systemic level [33]. Increasing adipose tissue mass leads to adverse changes in the profile of secreted adipokines as well as increased turnover of free fatty acids, which lead to increased insulin resistance. Interestingly, beneficial changes in glucose metabolism have been shown to appear soon after the initiation of an energy-restricted diet, even before any significant reduction in body fat is seen, which suggests that there are several different mechanisms in play simultaneously [34].

Chapter 6 provides a detailed description of the role of adiposity in glucose regulation.

Two of the most effective diabetes prevention studies (DPP and DPS) achieved an average of 5–7% weight loss among participants in the lifestyle intervention while post-hoc analysis from DPP identified weight loss as the main driver of changes in diabetes incidence, with each kilogram of weight loss being associated with a relative reduction of 16% in the risk of progression to T2DM [35]. Others have reported that weight reduction seems to be beneficial regardless of the mechanism of weight loss (e.g. diet or physical activity or both) [36], a hypothesis that is supported by studies showing significant reduction in the risk of T2DM following weight loss initiated through pharmaceutical intervention [37].

**Physical activity**

Whilst there is little doubt that weight loss is generally important to the prevention of T2DM, it is also true that changing other facets of lifestyle behavior can have profound effects on glucose regulation even in the absence of weight loss. Therefore, weight loss is not a prerequisite for improved glucose regulation. For example, the lifestyle intervention within the Indian Diabetes Prevention Programme did not result in weight loss despite a significant reduction in the relative risk of T2DM of around 30% compared to the control group [28]. This is an important message considering that the majority of individuals find weight reduction difficult to sustain in the long run, meaning other factors are needed if long-term reductions in the risk of T2DM are to be achieved. One of the most potent and universally applicable adiposity-independent therapeutic agents in the prevention of T2DM is physical activity.

Undertaking insufficient physical activity is associated with over 20 diseases and conditions and is estimated by the World Health Organization to be the fourth leading cause of premature mortality globally [38], ahead of both obesity and dietary factors. Somewhat rarely for lifestyle factors, evidence for a causal link between physical activity and the prevention of T2DM is supported by the full spectrum of evidence needed to infer causality, from observational research [39,40], to experimental mechanistic exploration [41–43] to randomized controlled trials and meta-analyses [24,47].

Observational research has confirmed that physical activity is a strong determinant of health in those with prediabetes. For example, in a large international cohort with prediabetes every 2000 steps per day change in walking activity over a 12-month period, as measured by a pedometer, was associated with an 8% difference in the risk of cardiovascular disease [39]. Others have shown that undertaking 150 minutes of leisure time walking activity per week was associated with a 60% reduction in the relative risk of diabetes [40].

Mechanistic studies have identified multiple pathways linking physical activity to insulin-stimulated glucose uptake [41–43]. Interestingly, muscular contractions are also known to induce glucose uptake through insulin independent pathways, which are likely to involve the upregulation of AMP-activated kinase [42]. Furthermore, muscle is an active endocrine organ and releases a plethora of myokines which are regulated by muscle fiber activation. For example, IL-6 is released from vigorously contracting muscle in quantities up to 100 times greater than resting values; this is turn has profound anti-inflammatory
### Table 38.2 Characteristics of lifestyle diabetes prevention programs

<table>
<thead>
<tr>
<th>Study</th>
<th>Country/study name</th>
<th>Sample size</th>
<th>Inclusion criteria</th>
<th>Interventions</th>
<th>Lifestyle intervention targets</th>
<th>Intervention duration</th>
<th>Risk reduction at end of intervention period (% reduction compared to controls)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pan et al., 1997 [30]</td>
<td>China /The Da Qing IGT and Diabetes Study</td>
<td>530 (283/247)</td>
<td>IGT, Age ≥25</td>
<td>1 Control 2 Diet 3 Exercise 4 Diet and exercise</td>
<td>Diet group: Weight maintenance for those with a BMI &lt;25 kg m(^{-2}) through a healthy isocaloric diet. Weight reduction for those with a BMI ≥25 kg m(^{-2}) through reduced energy intake&lt;br&gt;Exercise group: Participants were encouraged to increase the amount of their physical activity by at least one unit per day (such as slow walking for 30 min, or fast walking for 20 min) and by two units per day if possible&lt;br&gt;Lifestyle and diet group: Combination of above</td>
<td>6</td>
<td>Diet: 31&lt;br&gt;Exercise: 46&lt;br&gt;Exercise and diet: 42</td>
</tr>
<tr>
<td>Knowler et al., 2002 [27]</td>
<td>USA / Diabetes Prevention Research Group</td>
<td>3234 (1043/2191)</td>
<td>IGT, Age ≥25 yr, BMI ≥24 kg m(^{-2}) (≥22 kg m(^{-2}) if Asian), fasting plasma glucose ≥5.3 mmol L(^{-1})</td>
<td>1 Control 2 Lifestyle Metformin</td>
<td>150 min per week of moderate-to-vigorous physical activity and weight reduction (7% of initial body mass) through a healthy, low-energy, low-fat diet</td>
<td>2.8</td>
<td>58</td>
</tr>
<tr>
<td>Tuomilehto et al., 2001 [26]</td>
<td>Finland / Finish Diabetes Prevention Study</td>
<td>522 (172/350)</td>
<td>IGT, age 40 to 64 yrs old, BMI ≥25</td>
<td>1 Control 2 Lifestyle</td>
<td>30 min per day of moderate-to-vigorous physical activity and weight reduction (5% of initial body mass) through a healthy diet based on reduced saturated fat (10% of energy intake), reduced fat (30% of total energy intake), and high fiber</td>
<td>3</td>
<td>58</td>
</tr>
<tr>
<td>Ramachandran et al., 2006 [28]</td>
<td>India / Indian Diabetes Prevention Programme</td>
<td>531 (420/111)</td>
<td>IGT</td>
<td>1 Control 2 Lifestyle Metformin 3 Lifestyle plus Metformin</td>
<td>30 mins per day of moderate-to-vigorous physical activity and healthy diet based on reduced energy intake with fiber-rich foods low in refined carbohydrates and fats</td>
<td>3</td>
<td>Lifestyle: 29&lt;br&gt;Metformin: 28</td>
</tr>
<tr>
<td>Kosaka et al., 2005 [29]</td>
<td>Japan</td>
<td>458 (4580)</td>
<td>IGT</td>
<td>1 Control 2 Lifestyle</td>
<td>30 mins per day of moderate-to-vigorous physical activity and weight reduction (0.5–1 kg weight loss per month) through a healthy diet</td>
<td>4</td>
<td>67</td>
</tr>
</tbody>
</table>
effects which can act to augment improvements in insulin action [44]. Physical activity acts to favorably alter the distribution and composition of adiposity within the body. For example, increased physical activity is associated with reduced levels of hepatic fat, even without whole body weight loss, and recent evidence points to the role of physical activity in the promotion of brown adipose tissue [45,46].

Randomized controlled trials and meta-analyses have demonstrated that physical activity interventions result in improved glucose regulation and a reduced risk of T2DM in those with prediabetes [24,47]. For example, an exercise intervention in those with IGT was shown to reduce 2-hour post challenge glucose values by 1.3 mmol L$^{-1}$ over 12 months and reduce the risk of T2DM by a 60% over 24 months [47,48]. This is consistent with other studies and the results of a meta-analysis which demonstrated that physical activity interventions result in around a 50% reduction in the relative risk of type diabetes and are as effective as other multifactorial interventions that also promoted a healthy diet and weight loss [24].

As mentioned earlier, many of the benefits of increased physical activity are independent of overall body weight and adiposity. Observational research and interventions have consistently demonstrated substantial improvements in glucose regulation and reduced cardiovascular disease risk with higher levels of physical activity despite no significant change to body weight or waist circumference [39,40,42,47,49]. Therefore it is clear that the promotion of physical activity significantly improves glucose regulation, even in the absence of weight loss.

Successful diabetes prevention programs have typically promoted at least 150 minutes of moderate intensity physical activity per week, which is consistent with current international and national physical activity recommendations for health [50,51]. However, post-hoc analysis of diabetes prevention studies and epidemiologic evidence has consistently demonstrated a dose–response relationship between the amount of physical activity undertaken and the risk of T2DM, and recommendations for health now specifically state that greater benefits will accrue through achieving higher volumes of physical activity [50,51].

**Diet**

In addition to weight loss, diabetes prevention programs have targeted a reduction in total and saturated fat, with increased fiber intake. For example, DPS had three dairy goals: (1) less than 30% of energy intake derived from fat, (2) less than 10% of energy intake derived from saturated fat, and (3) at least 15 g of fiber per 1000 kcal [26]. Post-hoc analysis subsequently revealed that each one of these factors was related to the risk of T2DM, independent of body weight; compared to the lowest quartile, the risk of T2DM was approximately doubled in the highest quartile of total and saturated fat intake, and reduced by 60% in the highest quartile of fiber intake [52]. Other observational research studies and experimental interventions have consistently demonstrated the importance of these factors, particularly increased fiber and reduced saturated fat, in helping promote improved insulin sensitivity and a reduced risk of T2DM [53]. Plausible mechanisms linking these factors to insulin action have been proposed, including reduced ectopic fat, reduced low-grade inflammation, changes in cell membrane phospholipids, and enhanced intestinal peptide secretion [54,55].

As well as specific dietary factors, observational studies have suggested that overall dietary patterns are also an important consideration. For example, those consuming a “Westernized diet” (intake of refined grains, red and processed meat, sugar-sweetened beverages, high-fat dairy) have substantially higher risk (RR = 1.59) compared to those who do not, independently of body weight [56]. At the opposite end of the spectrum, those consuming a prudent diet (intake of whole-grain cereal, vegetables, legumes, low-fat dairy products, nuts, seeds, fish, and poultry), have a lower risk of T2DM [56]. A recent intervention further confirmed the importance of dietary patterns, where those assigned to a Mediterranean diet, characterized by high intake of vegetables, fruit, legumes, extra virgin olive oil, nuts, fish, whole grains, and red wine, decreased their risk of T2DM by 50%, without reductions to body weight [57].

**Pharmaceutical intervention**

An expanding range of pharmaceutical agents targeting β-cell function or insulin sensitivity have been tested in the prevention of T2DM over the last two decades; these can be broadly grouped as metformin, thiazolidinediones, α-glucosidase inhibitors, meglitides, angiotensin-converting enzyme inhibitors, angiotensin receptor blockers, and weight loss therapies. The primary studies investigating each of these classes of agents in the prevention of T2DM are discussed in the following sections and the main characteristics are displayed in Table 38.3. In addition, we highlight some new classes of drugs that have potential in the prevention of diabetes.

**Metformin**

Metformin is the most widely researched and best understood agent in the prevention of T2DM. It is classed as a biguanide and acts through suppressing hepatic glucose production and inhibiting free fatty acid production and oxidation, therefore reducing insulin resistance and promoting peripheral glucose uptake [58]. A recent meta-analysis demonstrated an average reduction in the risk of T2DM of 40% with metformin [59]. In DPP, metformin was associated with a 31% reduction in the incidence of T2DM at 3 years with an NNT of 13.9 [11]. Metformin was found to be most effective in younger adults and those with a higher BMI [11]. Metformin is known to have some gastrointestinal side effects (diarrhea, flatulence, nausea, and vomiting), however the therapy is generally well tolerated and these side effects are typically not treatment limiting. For example in DPP, metformin was associated with increased gastrointestinal symptoms (77.8 compared with 30.7 events/100 person-years with placebo), but adherence to metformin treatment was only slightly affected (77% with placebo vs. 72% with metformin took at least 80% of the prescribed dose) [11].
### Table 38.3: Characteristics of diabetes prevention programs evaluating pharmaceutical agents

<table>
<thead>
<tr>
<th>Study name</th>
<th>Country</th>
<th>Agent(s) tested</th>
<th>Study design</th>
<th>Sample size (male/female)</th>
<th>Inclusion criteria</th>
<th>Intervention duration</th>
<th>Risk reduction at end of intervention (% reduction compared to placebo)</th>
<th>Most common side effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diabetes Prevention Program [27]</td>
<td>USA</td>
<td>Metformin</td>
<td>RCT (placebo-controlled)</td>
<td>3234 (1043/2191)</td>
<td>IGT, BMI ≥ 24 kg m⁻² (≥ 22 kg m⁻² if Asian), fasting plasma glucose ≥ 5.3 mmol L⁻¹</td>
<td>2.8</td>
<td>31</td>
<td>Gastrointestinal symptoms</td>
</tr>
<tr>
<td>STOP-NIDDM [12]</td>
<td>International</td>
<td>Acarbose</td>
<td>RCT (placebo-controlled)</td>
<td>1368 (673/695)</td>
<td>IGT, fasting between 5.6 and 7.8 mmol L⁻¹</td>
<td>3.3</td>
<td>25</td>
<td>Gastrointestinal symptoms</td>
</tr>
<tr>
<td>DREAM [13,75]</td>
<td>International</td>
<td>Rosiglitazone, Ramipril</td>
<td>2 x 2 factorial (placebo-controlled)</td>
<td>5269 (2149/3120)</td>
<td>IGT and/or IGT without known CVD or renal impairment</td>
<td>3</td>
<td>Rosiglitazone: 69 Ramipril: no effect</td>
<td>Rosiglitazone: edema, congestive heart failure, weight gain Ramipril: cough</td>
</tr>
<tr>
<td>ACT-NOW [65]</td>
<td>USA</td>
<td>Pioglitazone</td>
<td>RCT (placebo-controlled)</td>
<td>602 (58% female)</td>
<td>IGT, fasting between 5.3 and 7.8 mmol L⁻¹, and at least one other risk factor</td>
<td>2.4</td>
<td>72</td>
<td>Edema, weight gain</td>
</tr>
<tr>
<td>NAVIGATOR [71,76]</td>
<td>International</td>
<td>Nateglinide, Valsartan</td>
<td>2 x 2 factorial (placebo-controlled)</td>
<td>9306 (4595/4711)</td>
<td>IGT with established cardiovascular disease or cardiovascular risk factors</td>
<td>5</td>
<td>Nateglinide: no effect Valsartan: 1.4</td>
<td>Nateglinide: hypoglycemia Valsartan: hypotension-related effects</td>
</tr>
<tr>
<td>XENDOS [37]</td>
<td>Sweden</td>
<td>Orlistat</td>
<td>RCT (placebo-controlled)</td>
<td>3305 (1810/1495)</td>
<td>Nondiabetic obese adults</td>
<td>4</td>
<td>37</td>
<td>Gastrointestinal symptoms</td>
</tr>
</tbody>
</table>
Lactic acidosis, the most severe side effect of biguanides in general, is also rare with metformin. Nonetheless, as with the treatment of T2DM, clinical caution is needed for patients with limited renal function or heart failure.

As well as the primary effect on glucose uptake, several relevant pleiotropic effects have been postulated for metformin, specifically the prevention of cardiovascular disease and cancer [60,61]. However, whilst each of these potential benefits has supporting observational and mechanistic evidence, data from placebo-controlled randomized trials is lacking or inconclusive.

Given its proven efficacy, the limited side effects and possible additional cardioprotective action, metformin is increasingly considered a suitable alternative or adjunct to lifestyle intervention strategies in the prevention of T2DM. For example, ADA recommends that metformin should be considered for those with both IGT and IFG and one additional factor [10]. NICE has also recently recommended that metformin should be considered when lifestyle intervention has been tried and found to be ineffective [62].

**Thiazolidinediones**

Thiazolidinediones (TZDs) are hepatic and peripheral insulin sensitizers which stimulate peroxisome proliferator-activated receptor-γ (PPAR-γ) activity including the promotion of adipogenesis and fatty acid uptake in peripheral but not visceral fat [63]. TZDs favorably alter adipokine profiles, particular increased adiponectin, which is associated with increased insulin sensitivity. In addition TZDs are also associated with β-cell preservation [64].

Two types of TZDs, rosiglitazone and pioglitazone, have been rigorously assessed and found to reduce the risk of T2DM by 60–70% over a 2.6- to 3-year period in those with prediabetes [13,65]. For example the DREAM trial, which included over 5000 participants, demonstrated a 60% reduction in the relative risk of T2DM in those randomized to receive rosiglitazone with a NNT of 7 [13]. In addition, rosiglitazone significantly increased regression to normoglycemia by 70–80% compared to placebo.

However, the impressive efficacy of TZDs in the prevention of T2DM is counterbalanced by serious side effects, which makes their use clinically inappropriate for this group. Trials have shown significant weight gain (2.6–7 kg) compared to placebo, which is unsurprising given TZDs directly promote adipogenesis. More worrying, TZDs are also associated with an increase risk of cardiovascular disease, including increased myocardial infarction and heart failure, and other adverse health effects [66,67].

**Alpha-glucosidase inhibitors**

α-Glucosidase inhibitors primarily act to delay carbohydrate absorption, therefore blunting the postprandial glucose response and reducing the toxic effect of elevated glucose concentrations on pancreatic β cells; in addition, α-glucosidase inhibitors directly improve insulin sensitivity in individuals with IGT [68]. In STOP-NIDDM, an international double-blind placebo-controlled trial involving 1368 randomized participants with prediabetes, acarbose was found to reduce the risk of T2DM by 25% over 3.3 years [12]. Other studies in those with prediabetes have been shown to have similar effects [69]. Acarbose has also been postulated to reduce the risk of cardiovascular disease which is supported by some experimental research and a meta-analysis of studies predominately undertaken in those with T2DM [70]. However, no published study has yet been specifically designed to assess this outcome.

**Meglitinides**

Nateglinide is a D-phenylalanine analogue, which acts directly on the β cells to stimulate a rapid short-duration secretion of insulin, thereby controlling postprandial hyperglycemia. NAVIGATOR, a multinational randomized 2×2 factorial placebo-controlled trial involving over 9000 participants, was designed to test the efficacy of nateglinide and valsartan at lowering incidence of T2DM and cardiovascular events in high-risk participants with prediabetes [71]. At 5 years, no significant difference was seen between placebo and nateglinide in these primary outcomes.

**Angiotensin-converting enzyme inhibitors and angiotensin receptor blockers**

In addition to cardiovascular benefits in those with T2DM, renin-angiotensin system (RAS) inhibitors have consistently been related to a reduced risk of T2DM in post-hoc analysis of antihypertensive trials [72–74]; however, these trials were limited by possible adverse effects to glucose control from the active comparators used. Only valsartan and ramipril have been specifically evaluated in the prevention of T2DM. In the DREAM trial, ramipril was not related to a reduced incidence of T2DM or CVD compared to placebo [75]. In the Navigator trial, valartan was associated with a small (14%) reduced risk of T2DM compared to placebo; however, no effect was seen on cardiovascular outcomes [76].

**Insulin analogues**

Given that β-cell dysfunction is one of the underlying causes of T2DM and is already present in prediabetes, early treatment with insulin analogue therapy may be effective at reducing the progression to T2DM in high-risk individuals. This hypothesis was recently assessed in the ORIGIN study; an international double-blind study with a 2×2 factorial design comparing insulin glargine and N-3 fatty acid supplementation against placebo [77]. In total, 12,537 patients with cardiovascular disease or at high risk for cardiovascular event who had prediabetes or diabetes were randomly assigned to receive insulin glargine, N-3 fatty acids supplementation or placebo. Among the 1456 patients with prediabetes at baseline, insulin glargine was associated with a 20–30% reduction in the relative risk of progressing to T2DM, depending on the time point assessed [77]. However, rates of weight gain and hypoglycemia were greater in the insulin glargine group and treatment was not
associated with a reduction in cardiovascular events after 6 years in the study cohort overall [77].

**Weight loss therapies**

Given the primacy of obesity in the development of T2DM, weight loss is an obvious therapeutic target for prevention. However, weight-reducing agents have only been studied in small numbers and only one, orlistat (an intestinal lipase inhibitor), has been specifically evaluated in the prevention of T2DM. In the XENDOS trial, 3304 nondiabetic obese subjects received lifestyle modification and were randomized to either orlistat or placebo for 4 years [37]. There was a weight loss of 6.9 kg in the orlistat group (compared with 4.1 kg in the placebo group) after 4 years that was associated with a 37% reduction in the relative risk of T2DM; this figure rose to 45% in those with IGT at baseline. However, low rate of completion (52% in the orlistat group) limits the conclusions that can be drawn from this study suggests poor acceptability among participants.

**Other agents**

There are several newer classes of therapies developed for use in the treatment of T2DM that target glucose control and weight loss, and show strong potential in the prevention of T2DM.

**Glucagon like peptide-1 (GLP-1) analogues and dipeptidyl peptidase-4 (DPP-4) inhibitors**

GLP-1 is a potent insulin secretagogue which is released by the L cells of the large intestine in response to food ingestion. GLP-1 analogues that mimic the action of GLP-1, but which are resistant to DPP-4 degradation, such as liraglutide and exenatide, have proven to be effective in the treatment of T2DM [78]. Similarly, therapies that inhibit the action of DPP-4, and hence slow the degradation of endogenous GLP-1, have also proven effective [78]. These therapies are attractive in the prevention of T2DM because they are glucose dependent, meaning their effect on insulin secretion is proportionate to the amount of circulating glucose, thus reducing the risk of hypoglycemia. In addition, GLP-1 analogues result in significant and sustained weight loss, with the greatest effects seen in those with the highest levels of obesity [79]. For example, just 20 weeks of liraglutide therapy has been shown to be more effective than orlistat in the treatment of obesity and resulted in a reduced prevalence of prediabetes by 84–96% depending on the dosage used [80]. Ongoing studies are testing the efficacy of GLP-1 therapy in the prevention of T2DM in high-risk groups.

**Sodium glucose co-transporter 2 (SGLT2) inhibitors**

SGLT2 inhibitors are a new class of therapy that work independently of insulin by preventing glucose reabsorption from glomerular filtration, resulting in glycosuria and a net loss of energy. SGLT2 inhibitors have been shown to improve glycemic control and induce weight loss of up to 5 kg resulting from an average energy loss of around 22–300 kcal per day [81]. However, studies have yet to be undertaken in the prevention of T2DM.

**Lifestyle versus pharmacotherapy**

Although national organizations and regulatory authorities are increasingly recommending the use of metformin, with the addition of other agents likely in the future, there remains some controversy around the use of pharmaceutical intervention as an adjunct, or in addition, to lifestyle intervention. First, few studies have assessed whether an additive benefit is gained by using lifestyle and pharmaceutical intervention in combination. The only study to have looked directly at outcome rates, the Indian Diabetes Prevention Programme, found that there was no additional benefit of combining metformin with lifestyle modification over and above lifestyle modification on its own in those with IGT [28]. Second, some drugs may actually act to attenuate the beneficial effects of lifestyle intervention. For example, there is mounting evidence that metformin may blunt the insulin sensitizing effects of exercise training in those with prediabetes or diabetes [82,83]. Third, lifestyle change is associated with a reduced risk of all-cause mortality, a reduced risk of cardiovascular morbidity and mortality, increased physical functioning and a reduced risk of depression [84]. Furthermore, an unhealthy lifestyle, not deficiency in metformin or other pharmaceutical agents, is the primary causal factor for T2DM. Therefore, lifestyle modification programs should be the focus of diabetes prevention initiatives. Whilst logistical and feasibility issues remain in implementing lifestyle programs in a primary healthcare or community settings (see following sections), it is the responsibility of funding bodies and research organizations to address this need. However, we acknowledge that pharmacotherapy may have a role to play when lifestyle modification programs have been tried and found to fail. This is consistent with new recommendations from NICE which advocate the use of metformin where high-risk individuals are unable to take part in a lifestyle intervention program or where lifestyle intervention has been ineffective at stabilizing or lowering glucose levels [62].

**Cost-effectiveness**

Cost-effectiveness modeling studies using different methods and assumptions have consistently found, based on results from the major prevention trials, that lifestyle intervention and metformin are likely to be cost-effective or incur a modest cost saving [85]. Only one study, out of over a dozen conducted to date, has found that lifestyle or metformin interventions may be near or outside the range considered cost-effective by most healthcare organizations (at $25,400 and $143,000 per QALY gained, respectively) [86]. Importantly, when considering the whole process from screening to treatment, Gillies et al. estimated that screening for T2DM and IGT and tailored treatment to each group were more cost-effective than screening for T2DM.
alone in the UK [87]; in addition the use of lifestyle intervention was more cost-effective than pharmaceutical therapy (£6242 vs. £7023 per QALY gained). NICE recently supported this work by suggesting that screening for T2DM and high-risk individuals would be cost-effective at £10,000 per QALY gained or less, and may be cost-saving for some groups [62].

Translation of diabetes prevention into routine primary care

National healthcare policies and funding structures have started to provide significant investment and commitment to the prevention of T2DM. This has necessitated a parallel shift in the focus of diabetes prevention research, to include the translation of effective prevention strategies into routine clinical practice [88]. This work has highlighted several key considerations around two specific areas: (1) suitable strategies to identify high-risk individuals, and (2) the incorporation of lifestyle interventions into routine clinical services.

Identifying those at high risk

Diabetes prevention initiatives have primarily focused on identifying and treating high-risk individuals with IGT, which is diagnosed through the oral glucose tolerance test (OGTT). However, there are important practical limitations regarding the utility and clinical value of carrying out OGTTs to identify those with a high risk of T2DM in routine care. Recently, this issue has been further complicated by the change in criteria for the diagnosis of T2DM to include HbA1c (at a threshold of 6.5% or higher) [17]; many healthcare organizations have, or are expected to, adopt HbA1c as the primary diagnostic tool for T2DM. It is therefore unlikely that the more invasive and burdensome OGTT would be acceptable in the identification of prediabetes. In addition, there are clinical concerns around patient compliance to fasting procedures and the test-retest reliability of 2-hour glucose values. Therefore, advocating the routine use of OGTTs as a screening tool in primary care is unlikely to be appropriate in most healthcare settings [54,89].

The utility of IGT and IFG have also been questioned in the prevention of T2DM [89]. There is high variation in the risk of developing both T2DM and cardiovascular disease across categories of prediabetes; importantly there is also a gradation of risk in those classified with normal glycaemia. For instance, data from DPS found that the risk of T2DM in those with IGT more than doubled with the presence of other readily identifiable risk factors [90]. It is also known that the risk of cardiovascular disease increases linearly with increasing levels of 2-h glucose and fasting glucose [91] and there does not appear to be a distinct threshold that justifies the use of distinct risk categories. Given these concerns, there has been much focus on developing pragmatic systematic approaches to the identification of individuals with an elevated risk of T2DM for referral into diabetes prevention initiatives. These have primarily evolved around the use of risk score technology.

Risk scores

Risk scores rely on utilizing clinically captured or self-assessed determinants or correlates of T2DM to calculate individual risk without the need for additional blood tests. Several risk assessment tools have been developed in the identification of those with a high risk of T2DM, the most widely validated and used of which is FINDRISC. FINDRISC was developed in Finland and uses weighted scores from eight risk characteristics to calculate an overall risk score [92]. FINDRISC has been shown to have good sensitivity (~0.8) and specificity (~0.8) at predicting the 10-year absolute risk of T2DM in a white European population. Similar results were seen for a risk score developed in Germany [93]. Further risk scores have been adapted to, and validated cross-sectionally in, diverse populations. For example, the Leicester Diabetes Risk Score, the Danish Diabetes Risk Score, and the Indian Diabetes Risk Score have all been associated with a sensitivity of 70–80% and a specificity of 55–75% [94–96] of detecting prediabetes and/or T2DM. These risk scores have typically been designed to be used by individuals to self-assess their own risk of T2DM; for example the Leicester Diabetes Risk Score was developed with input from patient groups in order to maximize uptake and comprehensibility [94]. However, recent innovations have included the development and validation of risk scores specially designed for use on data which is routinely available to and stored by general practitioners [97–99]. This allows practitioners to interrogate their patient databases to identify individuals with a high risk of T2DM. For example, the Leicester practice risk score, which uses six variables (sex, age, ethnicity, BMI, family history of T2DM, and antihypertensive medication) that are commonly coded within primary care databases in the UK, has been developed into an automated platform that automatically ranks all individuals registered within a given healthcare organization for their diabetes risk [98].

Two-stepped approach

Given their ease of use and noninvasive methodology, diabetes risk scores are now routinely used within many healthcare organizations. However, an emerging consensus (based on screening approaches used in clinical practice in the USA, Germany, Australia, Finland, the UK, and other countries) favors a targeted, staged approach whereby risk scores are used to identify moderate- to high-risk individuals and blood tests are then used to confirm risk status [54]. This approach has been shown to be the most cost-effective method of identifying those with a high risk of T2DM [100]. Following the most comprehensive review and in-depth analysis of the effectiveness and cost-effectiveness of different screening approaches to date, NICE developed a systematic two-stepped algorithm involving risk score technology to identify individuals ranking above the 50th percentile of risk; individuals meeting this criteria are then recommended to receive either a fasting or HbA1c blood
test to confirm their risk status before preventative action is undertaken [62].

Despite the potential clinical utility of diabetes risk score technology and simple blood tests in the identification of high-risk populations for referral into diabetes prevention initiatives, it is important to recognize that such approaches are not yet supported by gold-standard evidence. Previous diabetes prevention programs have involved specific populations identified on the basis of an OGTT. Therefore, there is a lack of data from randomized controlled trials investigating whether intervening in high-risk individuals, as identified through a risk score or a staged approach, is as effective at reducing the risk of developing T2DM. This is an important limitation because at-risk individuals identified through a risk score only or staged approaches may contain a relatively small proportion of individuals with IGT, the traditional target of intervention studies. Because IGT is specifically characterized by peripheral insulin resistance, it is likely to be more responsive to targeted intervention, particularly lifestyle, than other defined high-risk states.

**Implementing prevention programs**

As we have seen earlier, there is unequivocal evidence that lifestyle interventions work in the prevention of T2DM and are likely to be highly cost-effective. However, there is huge gap between evidence-based effectiveness and what is feasible in a real-world healthcare setting. Large clinical trials are conducted in ideal settings and apply significant and concentrated resources in pursuit of their goals [88]. For example, the lifestyle intervention within DPP involved 16 lengthy one-to-one counseling sessions, followed by monthly contact with in-person one-to-one contact at least once every 2 months and additional group-based sessions four times annually [11]. This degree of intervention intensity is far greater than even that offered to those with serious diseases in routine clinical care. Therefore, as others have noted, emphasis needs to be placed on examining the minimum level of intensity needed to produce meaningful clinical effects [101]. Beyond that, there are considerations around how new interventions become imbedded within usual care pathways and gain universal access. Prof. Ann Albright, Director for Division of Diabetes Translation within the Centers for Disease Control and Prevention (CDC), identified six distinct steps from basic science to national distribution, termed the continuum of translation, that have to be completed in order to achieve the universal implementation of diabetes prevention [102]. More recently, Schwarz and colleagues identified six key areas of focus to ensure sustainability when implementing diabetes prevention programs into practice, including: (1) intervention cost, (2) training and expertise of intervention providers, (3) uptake to both screening and intervention, (4) ensuring sustainability of funding and support within healthcare and political arenas, (5) developing quality management across intervention providers, and (6) using and improving technology to support behavior change of both patients and healthcare professionals [54].

Finland and the USA have been at the forefront of integrating diabetes prevention programs into usual healthcare systems. Finland adopted a regional systematic whole system approach across all sectors of the healthcare community and included population-based strategies, high-risk strategies, and early diagnosis and management strategies [103]. The high-risk strategy is based on using risk scores to screen the population in primary care, pharmacy, and community settings and then offering tailored lifestyle interventions based on the goals of DPP, generally delivered through 4–8 group-based sessions, to those found at high risk. In the USA, the Centers for Disease Control and Prevention (CDC) have received substantial funding and political support to implement a community-based diabetes prevention program run through YMCA and community facilities [104]. This program, consisting of 16 1-hour group-based sessions delivered by trained quality assured lifestyle coaches, has gained public and private support and has been integrated into the healthcare package offered by some health-insurance companies. Numerous smaller scale diabetes prevention programs, based on the goals of DPP, have also been developed and evaluated in the USA [105]. Internationally, European-wide diabetes prevention guidance and tools for healthcare professionals have also been developed and published [106].

Other groups have also recently begun to design and evaluate interventions that are specifically tailored to the structure and resources available to their national healthcare services. In the United Kingdom, a less resource-intensive, 3-hour group-based structured education program, based on models and systems used in the management T2DM, has proven highly effective at promoting health behavior change and improving glucose regulation in those with prediabetes over 1 year with results sustained at 2 years [47,48]. This theory-driven program, titled “Walking Away from Type 2 Diabetes,” has been fully developed into a nationally available diabetes prevention program with accredited educator training and quality assurance professional pathways [107]. Other similar approaches have been developed in Germany and Australia [88].

Intervention designed to translate the findings from diabetes prevention trials into practice, although developed concurrently across diverse healthcare settings and populations, have commonality across several key areas: they have concentrated on lifestyle, tended to utilize group-based education and counseling approaches, and have been delivered across a range of primary care and community venues by personnel largely outside of traditional clinical practice (i.e., other than general practitioners or nurses) [89,105,108]. Several recent systemic reviews and meta-analyses have found that where evaluated, these interventions continue to demonstrate positive outcomes on body weight and other markers of health behavior [105,108]. For example, the average weight loss in interventions modeled on DPP was 4 kg at 12 months compared to 6.5 kg seen in the original DPP trial. Therefore, whilst interventions aimed at translating diabetes prevention programs into routine care have, by necessity, had to use less...
resource-intensive methodology, they have continued to show sustained benefits.

High risk versus population

The focus of this chapter and the main discourse on diabetes prevention internationally has been on developing and evaluating upstream strategies designed for high-risk populations. By their very nature such approaches only treat the end-products of a deleterious system, they do not seek to identify and address the underlying fault. This was recognized by Geoffrey Rose 30 years ago with his widely publicized “prevention paradox,” which states that strategies designed to shift the population distribution of known risk factors are inherently the only way of effectively tackling a mass disease [109]. Some causal determinants of T2DM, such as obesity and physical inactivity, are now so common that only population-based interventions will begin to fully tackle these issues. This concept was highlighted in a recent study which estimated that a 1% decrease in BMI across the whole population (roughly equal to weight loss of 1 kg per person) would avoid 179,000 – 202,000 incident cases of diabetes, 122,000 cardiovascular diseases, and 32,000 – 33,000 incident cases of cancer over the next 20 years in the UK [110]. Therefore, along with high-risk and disease management strategies that remain important for widely prevalent conditions, mass action, similar to that employed to target smoking, is needed involving government policy, legislative and taxation support, and industry compliance to alter the default patterns of behavior associated with our everyday environments. Although infrequently attempted or evaluated, such downstream initiatives have been found to be effective at reducing cardiovascular risk factors within the population [111]. However, the “paradox” remains that interventions applied to the population can lead to little individual benefit and can even cause some individual dissatisfaction. For example, a healthy individual may wonder why their “right” to eat what they like is being infringed by a differential taxation system that targets certain food types. For this reason, employing such population-wide initiatives carries greater controversy and political risk and is therefore generally less likely to be enacted to the extent needed. However, to do nothing, or to continually be at the mercy of shifting and ineffectual political priorities, will expose millions of individuals to obesity and its consequences for generations to come.

Conclusion

It is clear that the prevention of T2DM is a public health priority. There is unequivocal evidence that lifestyle interventions, based on the promotion of physical activity, diet and weight loss, and some pharmaceutical interventions, are highly effective and cost-effective in the prevention of T2DM. Effective translation of lifestyle interventions into routine healthcare settings has remained a challenge; however, recent advances have demonstrated considerable success in effectively tailoring and delivering pragmatic diabetes prevention programs in “real world” routine clinical and community settings. These advances have been matched by government policies that are beginning to prioritize the prevention of chronic disease. Therefore, we stand at a unique crossroads; all the elements required to radically address and reverse this highly prevalent chronic disease have been developed. Whether or not these elements can be adequately combined to realize this potential will depend on whether the academic community, medical profession, healthcare organizations, local authorities, national governments, and industry can join forces and act together with a common purpose.

References

13 Gerstein HC, Yusuf S, Bosch J, et al.; DREAM (Diabetes REduction Assessment with ramipril and rosiglitazone Medication) trial: Effect of rosiglitazone on the frequency of diabetes in patients
with impaired glucose tolerance or impaired fasting glucose: a randomised controlled trial. *Lancet* 2006;368:1096–1105.


562 Chapter 38


Prevention of type 2 diabetes


105 Ali MK, Echouffo-Tcheugui J, Williamson DF: How effective were lifestyle interventions in real-world settings that were modeled on the diabetes prevention program? *Health Affairs* 2012;31:67–75.


CHAPTER 39
Prevention of diabetic microvascular complications

Alice P.S. Kong and Juliana C.N. Chan
Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong SAR, China; Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Hong Kong SAR, China

Key points
• Good control of blood glucose, blood pressure, and lipids are all vital in preventing diabetic microvascular complications. Structured care and an all-encompassing approach in intensified management of multiple risk factors are recommended.
• Renin-angiotensin-aldosterone system (RAAS) blockade with either angiotensin converting enzyme inhibitor (ACEI) or angiotensin II-receptor blocker (ARB) confers additional protective effects in primary and secondary prevention of diabetic nephropathy, a beneficial effect independent of blood pressure lowering.
• The role of other approaches to achieve complete blockade of RAAS including dual blockade with ACEI and ARB, aldosterone receptor antagonists or direct renin inhibitor need to be explored with more definitive clinical trial evidence.
• Therapies targeting multiple metabolic pathways in preventing the development of diabetes-related microvascular complications including diabetic neuropathy have been attempted and may shed light on future treatment.
• Antivascular endothelial growth factors (VEGF) inhibitors and anti-inflammatory agents such as corticosteroids may play a role in preventing the progression of diabetic retinopathy.

Introduction
Type 2 diabetes (T2DM) is a silent disease with devastating consequences. With increasing earlier onset of T2DM across the globe, particularly in Asia, the younger ages at which T2DM is diagnosed translates into an extended disease duration and an escalating rate of complications [1–3]. Life expectancy of diabetic patients is generally shorter than their nondiabetic counterparts [4], because diabetes is a well-established, independent risk factor for a wide range of vascular diseases, as well as nonvascular diseases, including cancers [4,5].

Macrovascular complications associated with diabetes are the leading cause of death among patients with diabetes [6], and diabetes confers about a twofold excess risk for coronary heart disease and ischemic stroke [5]. However, diabetes-related microvascular complications, namely nephropathy, retinopathy, and neuropathy, account for the excess morbidity, which negatively impacts on patients’ quality of life and poses a huge burden on the health system and resources [6,7]. Moreover, experiencing microvascular complications is associated with a greater risk of macrovascular complications, including coronary heart disease and cerebrovascular diseases [7,8]. Since prevention is more cost-effective than the treatment of diabetes complications, there is indeed a pressing need for more evidence to support efficacious preventative strategies that could delay or avert the development and progression of microvascular complications in diabetic populations. Here, we review the evidence from clinical trials in the prevention of diabetic complications, with a particular focus on the preventive strategies of diabetic microvascular complications.

Diabetic nephropathy
Diabetes is the single leading cause of end-stage renal disease worldwide and accounts for approximately 40% of patients requiring dialysis [9]. Due to the diversity of genetic and environmental factors, the prevalence and incidence of diabetic nephropathy varies across different ethnic groups. Diabetic kidney disease has emerged as an important cause of morbidity and mortality in diabetes, particularly in Asia where the prevalence of diabetic nephropathy is alarmingly high (up to 60%) when compared to the United States and Europe [1,10]. Diabetic nephropathy is particularly prevalent in Asian diabetic populations (40% have microalbuminuria and 20%
have macroalbuminuria) [10], and occurs in 20–40% of diabetic patients in the United States [11]. Diabetic patients with nephropathy and reduced renal function, as measured by estimated glomerular filtration rate (eGFR), are especially vulnerable to developing cardiovascular events and all-cause mortality [12].

**Intensive glycemic control in the prevention of diabetic nephropathy**

It has been well recognized that good glycemic control is crucial to prevent diabetic nephropathy in both type 1 and type 2 diabetic patients. The Diabetes Control and Complications Trial (DCCT) showed that tight glycemic control, close to normoglycemia, delays the onset and slows the progression of diabetic nephropathy, retinopathy, and neuropathy in type 1 diabetes [13]. The DCCT randomly assigned 1441 patients with type 1 diabetes (mean age 27 ± 7 years) to intensive or conventional therapy, treating them for a mean of 6.5 years between 1983 and 1993 [13]. After a mean follow-up of 17 years (93% were followed until February 2005 during the observational Epidemiology of Diabetes Interventions and Complications study (EDIC)), type 1 diabetic patients in the intensive treatment group had a significantly lower risk of experiencing diabetic nephropathy (microalbuminuria and albuminuria), cardiovascular and myocardial infarction compared to those in the conventional therapy group [14].

In keeping with the findings of the DCCT, the United Kingdom Prospective Diabetes Study (UKPDS) [15] (a 10-year randomized study in the UK including 3867 newly diagnosed type 2 diabetic patients) and the Kumamoto Study [16] (a 6-year randomized study in Japan including 110 type 2 diabetic patients with poor glycemic control), also showed that intensive glycemic control reduces the risk of microvascular complications, including diabetic nephropathy [15,16]. In the Kumamoto study, 110 Japanese type 2 diabetic patients with poor glycemic control were randomized to either the multiple insulin injection group or the conventional insulin injection group to evaluate the impact of intensive treatment versus conventional treatment, on both primary and secondary prevention of diabetic nephropathy, retinopathy, and neuropathy (mean glycated hemoglobin (HbA1c) at baseline was 8.9 ± 1.8% and 9.2 ± 1.8% in the primary prevention group; 9.0 ± 1.9% and 9.4 ± 1.8% in the secondary prevention group) [16]. After 6 years of treatment, diabetic patients in the intensive treatment group (mean HbA1c 7.1 ± 1.1%) had a significantly reduced risk of suffering the onset and progression of diabetic nephropathy, as well as other diabetes-related microvascular complications, namely retinopathy and neuropathy, compared to those in the conventional treatment group (mean HbA1c 9.4 ± 1.5%) [16]. The cumulative percentage of the development and the progression in nephropathy in the primary prevention cohort and secondary prevention cohort were 7.7% and 28.0%, and 11.5% and 32.0% in the intensive and conventional group, respectively [16]. In the UKPDS, 3867 newly diagnosed type 2 diabetic patients (median age 54 years, interquartile range 48–60 years) were randomly assigned to the intensive treatment or conventional treatment group. After 10 years, diabetes patients in the intensive group (mean HbA1c 7%) had a 25% risk reduction in the development of microvascular complications compared to the diabetic patients in the conventional group (mean HbA1c 7.9%) [15]. The 10-year post-trial follow-up of the UKPDS [17], similar to the long-term follow-up data of the DCCT [14] and the Kumamoto study [18], demonstrated the beneficial effect regarding the legacy of intensive glycemic control in delaying the onset and progression of microvascular complications.

Since then, a number of randomized controlled clinical trials, including the ADVANCE trial (Action in Diabetes and Vascular Disease: Preterax and Diamicron Modified Release Controlled Evaluation) [19], ACCORD (Action to Control Cardiovascular Risk in Diabetes) [20], VADT (Veteran Affairs Diabetes Trial) [21], and the VA (Veterans Affairs) Feasibility Trial [22] were conducted, which consistently reaffirm that intensive glucose control reduces the risk of diabetic nephropathy among type 2 diabetic patients (Table 39.1). In the ADVANCE study [19], 11,140 type 2 diabetic patients with poor glycemic control (median HbA1c at baseline 7.2%) were recruited from 20 countries spanning Asia (37%), Australasia (13%), Europe (46%), and North America (4%). The study subjects had a mean disease duration of 8 years and a mean age of 66 years. The eligible subjects were randomly assigned to the intensive treatment group, using glinides (moderate release) plus other glucose-lowering drugs aiming to keep HbA1c <6.5%, or to the conventional treatment group with an HbA1c goal according to the local guidelines. After a median follow-up of 5 years, the mean HbA1c was lower in the intensive group than in the conventional treatment group (6.5% vs. 7.3%), and the intensive group was associated with a 21% reduction in renal events including new or worsening nephropathy (p = 0.006) and a 9% reduction of new-onset microalbuminuria (p = 0.02) [19]. In the ACCORD study [20], 10,251 type 2 diabetic patients were recruited in the United States and Canada (mean age of 62 years, mean disease duration of 10 years, median HbA1c at baseline 8.1% with the goal of HbA1c for the intensive arm of <6% vs. 7–7.9% for the conventional group). The study was prematurely terminated after 3.7 years of treatment due to a higher mortality rate in the intensive treatment group. The results of the ACCORD study showed that there was fewer renal events in the intensive glycemic arm [20]. In comparison, the VADT [21] and VA Feasibility Trial [22] were on a smaller scale with 1791 and 153 male patients with T2DM studied, respectively. The results from both the VADT [21] and VA Feasibility Trial [22] support the finding that intensive glycemic control reduces the risk of diabetic nephropathy compared to conventional treatment.
### Table 39.1

Randomized controlled trials examining intensive glycemic control and risk of microvascular complications in diabetic patients.

Legend: ↓ decreased risk; ↔ no change in risk; white: initial study; grey: follow-up study. DCCT: The Diabetes Control and Complications Trial [13]; EDIC: Epidemiology of Diabetes Interventions and Complications study [14]; UKPDS: United Kingdom Prospective Diabetes Study initial study [15] and follow-up study [17]; Kumamoto Study [16]; ADVANCE: Action in Diabetes and Vascular Disease: Preterax and Diamicron Modified Release Controlled Evaluation [19]; ACCORD: Action to Control Cardiovascular Risk in Diabetes [20]; VADT: Veteran Affairs Diabetes Trial [21], and VA (Veterans Affairs) Feasibility Trial [22]

<table>
<thead>
<tr>
<th>Study</th>
<th>Diabetic Patients</th>
<th>Diabetic Nephropathy</th>
<th>Diabetic Retinopathy</th>
<th>Diabetic Neuropathy</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCCT/EDIC</td>
<td>1441 type 1 diabetes patients (mean age 27±7 years, mean disease duration around 2 to 3 years in primary prevention cohort and around 8 to 9 years in secondary prevention cohort) and mean duration of study of 6.5 years in the initial study; post-trial follow-up ranged from 6.5 to 17 years</td>
<td>↓</td>
<td>↓</td>
<td>↓</td>
</tr>
<tr>
<td>UKPDS</td>
<td>3867 newly diagnosed type 2 diabetic patients (median age 54 years, interquartile range 48–60 years) studied over 10 years and another 10 years post-trial follow-up</td>
<td>↓</td>
<td>↓</td>
<td>↔ No data</td>
</tr>
<tr>
<td>Kumamoto</td>
<td>110 Japanese type 2 diabetic patients (mean age around 47–52 years, mean disease duration around 6–7 years in primary cohort and around 10 years in secondary prevention cohort) for a treatment period of 6 years and another 8 years of post-trial follow-up</td>
<td>↓</td>
<td>↓</td>
<td>↓</td>
</tr>
<tr>
<td>ADVANCE</td>
<td>11,140 type 2 diabetic patients (mean age of 66 years, mean disease duration of 8 years, 46% from Europe &amp; 37% from Asia, median HbA1c at baseline was 7.2%, goal of HbA1c for intensive arm ≤6.5%, median follow-up of 5 years)</td>
<td>↓</td>
<td>↔</td>
<td>↔</td>
</tr>
<tr>
<td>ACCORD</td>
<td>10,251 type 2 diabetic patients in United States and Canada (mean age of 62 years, mean disease duration of 10 years, median HbA1c at baseline was 8.1%, goal of HbA1c for intensive arm &lt;6%, premature termination after 3.7 years of treatment due to higher mortality in the intensive treatment group)</td>
<td>↓</td>
<td>↓</td>
<td>↓</td>
</tr>
<tr>
<td>VADT</td>
<td>1,791 type 2 diabetic patients from United States (mean age of 60 years, mean disease duration of 11.5 years, median HbA1c at baseline was 9.4%, goal of HbA1c for intensive arm &lt;6%, median duration of follow-up of 5.6 years)</td>
<td>↓</td>
<td>↔</td>
<td>↔</td>
</tr>
<tr>
<td>VA Feasibility Trial</td>
<td>153 type 2 diabetic patients (mean only, mean age of 60 years, mean disease duration of 8 years) were treated for 2 years</td>
<td>↓</td>
<td>No data</td>
<td>No data</td>
</tr>
</tbody>
</table>

|   | Initial Study | Follow up study | ↓ Decreased risks | ↔ No increased or decreased risks |

**Blood pressure control, renin-angiotension-aldosterone system and prevention of diabetic nephropathy**

Apart from intensive glycemic control, achieving tight control of blood pressure and blockade of the renin-angiotensin-aldosterone system (RAAS) are the cornerstones in preventative strategies of diabetic kidney disease, as demonstrated by a number of clinical trials [23].

In addition to demonstrating the beneficial effect of tight glycemic control, the UKPDS also clearly shows that tight blood pressure control (aiming at blood pressure <150/85 mmHg vs. <180/105 mmHg with either angiotensin-converting enzyme inhibitor or a beta-blocker) reduces the risk of developing microvascular complications by 37% [24]. For every 10 mmHg reduction in systolic blood pressure, there is a 12% risk reduction in any diabetes-related complications [25]. The beneficial effect of blood pressure control is further supported a number of clinical trials, including the RENAAL (Reduction of Endpoints in NIDDM with the Angiotensin II Antagonist Losartan) study [26] which showed that every 10 mmHg increase in baseline systolic blood pressure was associated with an increased risk of end-stage renal disease (ESRD) or death by 6.7%, after adjusting for potential confounders including urinary ACR, serum creatinine, serum albumin, hemoglobin, and HbA1c [27].

There is little doubt about the role of RAAS blockade in conferring protective effects in patients with diabetic nephropathy, which is independent of blood pressure lowering. The IDNT (Irbesartan Diabetic Nephropathy Trial) [28] and RENAAL [26] study are the landmark trials that clearly demonstrated the renoprotective effects of both angiotensin-converting enzyme inhibitor (ACEI) and angiotensin II-receptor blocker (ARB). In the IDNT study, a total of 1715 patients with hypertension...
and diabetic nephropathy were randomized to either irbesartan (300 mg daily) or placebo (mean duration of follow-up was 2.6 years) [28]. Irbesartan treatment was associated with a 33% risk reduction in doubling of serum creatinine and a 23% risk reduction in ESRD, when compared to the placebo group [28]. In the RENAAL study, a total of 1513 patients with T2DM and diabetic nephropathy were randomized to either losartan (50 to 100 mg daily) or placebo [26]. After a mean treatment of 3.4 years, patients in the losartan group had a 25% risk reduction in the incidence of doubling of serum creatinine level and 28% risk reduction in the incidence of ESRD, compared to the placebo group [26]. The subgroup analysis of the RENAAL study also showed that the losartan conferred an even greater risk reduction in Asian diabetic patients with a 35% risk reduction in the primary composite endpoint composed of a doubling of serum creatinine, ESRD, and all-cause mortality [29].

Regarding the comparative effectiveness of ACEI versus ARB in secondary prevention of renal deterioration in patients with diabetic nephropathy, data from the DETAIL (Diabetics Exposed to Telmisartan and Enalapril) study did not support any advantage of one over the other [30]. Recently, the ROADMAP (Randomized Olmesartan and Diabetes Microalbuminuria Prevention) trial showed clinical evidence of ARB in the primary prevention of diabetic nephropathy. In ROADMAP, olmesartan, an angiotensin II-receptor blocker (ARB), is shown to delay onset of microalbuminuria in patients with T2DM [31]. The ROADMAP study included 4447 type 2 diabetic patients (mean age 57.7 years, 46.1% men) randomly assigned to receive olmesartan or placebo for a median duration of 3.2 years. There was a 23% risk reduction of development of microalbuminuria in type 2 diabetic patients in the olmesartan group [31].

Despite optimal control of glycemia, blood pressure, lipids and the use of RAAS blockers, the renal function of some diabetic patients with nephropathy continues to deteriorate and not all patients respond to treatment of ACEI or ARB. Since 40% of angiotensin II is produced by other non-ACE pathways, such as chymase, which may be substantially more active in diabetic patients [32], this incomplete blockade of RAAS may explain the observation that some diabetic patients continue to have rapid renal deterioration despite treatment with ACEI or ARB, and in addition to those treatment nonresponders. Theoretically, treatment with ARB may result in a more complete blockade of the RAAS, and dual blockade of the RAAS with both ACEI and ARB has been advocated to achieve better renal protection. However, the result of the ONTARGET (Ongoing Telmisartan Along and in Combination with Ramipril Global Endpoint Trial) [33] is disappointing with a worsening of renal outcomes after dual blockade of RAAS. ONTARGET is a mega trial which included 25,620 patients with established atherosclerotic vascular disease or with diabetes with end-organ damage and randomized them to ramipril (n = 8576), telmisartan (n = 8542), or a combination of both drugs (n = 8502) with a median follow-up of 56 months [33]. The event rate for primary renal outcome (dialysis, doubling of serum creatinine and death) was similar for the telmisartan and ramipril groups but was increased in patients treated with dual blockade of ACEI and ARB [33]. The ORIENT (Omesartan Reducing Incidence of Endstage Renal Disease in Diabetic Nephropathy Trial) [34] also failed to show that the use of ARB in combination with ACEI improves renal outcomes in patients with diabetic nephropathy. ORIENT recruited Asian type 2 diabetic patients (377 Japanese and 200 Chinese) with overt diabetic nephropathy, who were either on (73.5% in total cohort and 100% in Chinese participants) or not on baseline ACEI treatment, to receive either olmesartan (n = 288) or placebo (n = 289) over a mean treatment duration of 3.2 years [34]. The primary composite outcome (doubling of serum creatinine, endstage renal disease and death) was not significantly different in the olmesartan and placebo groups [34]. Of note, the predefined secondary cardiovascular endpoint was positive in ORIENT study, similar to a study involving Valsartan or another one involving spironolactone which showed that dual blockade was cardioprotective.

Other new approaches to achieve complete blockade of RAAS have been attempted in the treatment of patients with diabetic nephropathy. Aldosterone receptor antagonists, such as spironolactone, have been shown to further reduce proteinuria when added to an ACEI or ARB in treating diabetic patients with nephropathy. However, only small-scale and short-term clinical trial evidence is available [35,36] and thus long-term clinical benefits are yet to be shown.

Aliskiren, the first direct renin inhibitor that is orally active, has been approved for the treatment of hypertension since 2007. Aliskiren differs from ACEI and ARB in blocking the first and rate-limiting step in the RAAS, leading to a suppression of all RAAS hormones, including plasma renin activity but not plasma renin concentration. Therefore, aliskiren has the theoretical advantage of overcoming incomplete RAAS blockade with ACEI and ARB, and may translate into improved clinical benefits in treating patients with diabetic nephropathy. In the AVOID (Aliskiren in the Evaluation of Proteinuria in Diabetes) study [37], 599 type 2 diabetic patients with nephropathy, hypertension, and treated with 100 mg daily losartan at baseline, were randomized to receive 6 months therapy of add-on aliskiren or placebo. Patients in the aliskiren group had a 20% reduction of urinary ACR compared to the placebo group [37]. Following this encouraging preliminary data from AVOID, the ALTITUDE (Aliskiren Trial in Type 2 Diabetes Using Cardio-Renal Endpoints) [38] trial was designed to examine whether dual RAAS blockade with aliskiren in combination with an ACEI or ARB in patients with type 2 diabetic nephropathy and renal impairment would reduce cardio renal endpoints. To the dismay of many clinicians and researchers, ALTITUDE was prematurely terminated due to the increased incidence of nonfatal stroke, renal complications, hyperkalemia, and hypotension after 18–24 months of follow-up [39]. Aliskiren is subsequently not recommended by regulatory authorities for prescription with ACEI and ARB in patients with T2DM and renal impairment. Pending disclosure of the full results, the early termination of
the ALTITUDE trial further highlights the concerns regarding the safety issues of complete blockade of RAAS.

**Other therapies for preventing diabetic nephropathy**

Dyslipidemia is known to be associated with diabetic vascular complications including nephropathy. In a sub-study of the FIELD (Fenofibrate Intervention and Event Lowering in Diabetes) study [40], reduced albuminuria and a slower decline of eGFR over 5 years were observed, suggesting fenofibrate may prevent the onset and progression of diabetic nephropathy in type 2 diabetic patients, despite an initial increase in plasma creatinine levels, which was reversible [41]. The FIELD study was a multinational study including 9795 type 2 diabetic patients (mostly White/Caucasian) aged 50–75 years (mean age 62 years) randomly assigned to receive fenofibrate 200 mg daily or placebo treatment for an average of 5 years, with the aim to assess whether lipid-lowering therapy could reduce macro- and microvascular complications in type 2 diabetic patients [40].

Since the pathogenesis of diabetic nephropathy is known to be associated with abnormalities of renal nitric oxide (NO) generation, inhibition of phosphodiesterase 5 (PDE5) may have a renoprotective potential in patients with diabetic nephropathy. PDE5 is an enzyme that degrades cyclic guanosine monophosphate (cGMP), an intracellular mediator generated following the activation of soluble guanylate cyclase by NO. The inhibition of PDE5 thus leads to the elevation of intracellular cGMP levels. Many of the biologic actions of NO are mediated by cGMP, including the relaxation of vascular smooth muscle. Data from animal models showed that sildenafil, a PDE5 inhibitor, decreased albuminuria, attenuated glomerular hyperfiltration and resulted in a reduction of glomerular hypertrophy and glomerulosclerosis score, and ameliorated oxidative and inflammatory renal injuries in type 2 diabetic rats [42,43].

More novel therapies have recently begun to emerge, with the vast majority focusing upon the anti-inflammatory pathways, as well as kidney regeneration, stem cell growth factors, and so on. Yet, these novel therapies are still at the infancy stage of development, but may help to shed light on and hope for future treatments for T2DM.

Given that diabetes is a multifaceted disease, an intensified multifactorial intervention, incorporating a combination of tightly controlled glycemia, optimal blood pressure control, the use of RAAS blocker, and lipid-lowering agents, plays a pivotal role in the primary and secondary prevention of diabetic nephropathy. The STENO study clearly demonstrated the importance of this all-encompassing approach in intensified management of risk factors in treating type 2 diabetic patients [44]. In the STENO-2 study [44], 160 type 2 diabetic patients with microalbuminuria were randomized to receive either intensive therapy (multiple drugs for tight glycemia, blood pressure, and lipid control as well as aspirin) or conventional therapy. The mean treatment of 7.8 years was followed by a mean observational period of 5.5 years. Diabetic patients in the intensive treatment group experienced a reduced progression to endstage renal disease and required less retinal photocoagulation compared to the conventional group [44]. The clinical benefit of a structured care model in reducing renal endpoints is further supported by the SURE study [45] (Figure 39.1).

**Diabetic neuropathy**

Diabetic neuropathy is amongst the most common of the long-term complications of diabetes affecting up to 50% of patients [46]. The epidemiology and natural history of diabetic neuropathy, both somatic and autonomic, remain poorly defined, partly because existing studies use different and variable criteria to diagnose diabetic neuropathy. Diabetic neuropathy, in combination with the occurrence of other complications, such as peripheral vascular disease, can result in diabetic foot disease and is the leading cause of amputation.

**Role of metabolic and blood pressure control in diabetic neuropathy**

The DCCT clearly demonstrates the beneficial effects of intensive glycemic control in delaying the onset and progression of diabetic neuropathy in T1DM [13]. The long-term follow-up in the EDIC trial of subjects originally from the DCCT cohort, supports the unique legacy effect of prior tightly controlled blood glucose in delaying the onset of incident neuropathy [47] including cardiac autonomic neuropathy [48], an important cause of sudden death and one of the most overlooked complications of diabetic patients.

In the UKPDS, there was no significant difference in diabetic neuropathy, as assessed clinically by knee (11 vs. 12%, p = 0.42) and ankle reflexes (35 vs. 37%, p = 0.60), between the intensive and conventional treatment groups [17]; autonomic neuropathy was also assessed by R–R intervals measured on electrocardiograms at expiration and inspiration on deep breathing for five cycles, change in R–R interval on standing, and, in men, self-reported erectile dysfunction [15]. Likewise, the incidence of new or worsening neuropathy was similar in type 2 diabetic patients in both the intensive and standard treatment groups in the ADVANCE study [19]. By contrast, results from the Kumamoto [16] and ACCORD [20] studies support the role of tight glycemic control in the prevention of diabetic neuropathy (Table 39.1). In the Kumamoto study, intensive glycemic control was shown, in both the initial study and long-term follow-up, to delay the onset and impede the progression of diabetic neuropathy, diabetic nephropathy, and retinopathy [18,49]. Regarding blood pressure control, although the ABCD trial (Appropriate Blood Pressure Control on Diabetes) [50] failed to show intensive blood pressure control reduced diabetic neuropathy or retinopathy compared to moderate blood pressure control (a 5-year study including 470 hypertensive diabetes patients with a goal to maintain diastolic blood pressure ≤75 mmHg vs. 80–89 mmHg), the STENO-2 study showed that the progression of autonomic neuropathy, as measured by the R–R
Prevention of diabetic microvascular complications

Figure 39.1 The SURE (Effects of Structured versus Usual Care on Renal Endpoints in type 2 diabetes) study [45] showed a significantly higher percentage of diabetic patients in the structured care group achieved at least 3 predefined treatment targets compared to diabetic patients randomized to the usual care group (61% vs. 28%, p < 0.001) (a). Treatment targets were defined as HbA1c < 7%, blood pressure < 130/80 mmHg, LDL cholesterol < 2.6 mmol L\(^{-1}\), triglyceride < 2 mmol L\(^{-1}\), and persistent treatment with renin-angiotensin blockers. Patients who attained ≥3 treatment targets had reduced risk of primary endpoint (death and/or renal endpoint, i.e., serum creatinine > 500 μmol L\(^{-1}\) or dialysis) compared to those who attained ≤2 targets (relative risk 0.43; 95% CI 0.21–0.86) (b).

Strategies in preventing diabetic neuropathy: from pathophysiology to prevention

The underlying mechanisms leading to the development of diabetic neuropathy are still not completely understood [46]. Nonetheless, there is a large body of evidence suggesting that multiple metabolic pathways may interact and contribute to nerve damage [46]. Hyperglycemia activates several metabolic processes including the polyol pathway, oxidative stress, protein kinase C-beta (PKC-β), formation of advanced glycation endproducts (AGEs), which in turn act on specific receptors (RAGES) inducing monocytes and endothelial cells to increase production of cytokines and adhesion molecules [46]. Moreover, growth factors (e.g. ciliary neuropathic factor, insulin-like growth factors, vascular endothelial growth factors) and autoimmune immunoglobulin may also play a role in the prevention of
diabetic neuropathy in in vitro studies. Therapies targeting these pathways have been attempted in animal models and small-scale human studies, but evidence from clinical trials to support the use of these novel agents is still lacking. For instance, aldose reductase is the key enzyme in the polyol pathway, and the blockade of this enzyme prevents the generation of neurotoxic sugar alcohols, for example sorbitol. Aldose reductase inhibitors have been studied in a number of preliminary trials, but many had significant methodological flaws [51]. From a meta-analysis including thirty-two randomized controlled trials, there was no significant difference between aldose reductase inhibitors and placebo in the treatment of diabetic polyneuropathy [51]. However, in another meta-analysis including ten trials which aimed to explore the role of aldose reductase inhibitors in the treatment of diabetic cardiovascular autonomic neuropathy, aldose reductase inhibitors were effective especially in mild or asymptomatic cases with minimal adverse effects except Tolerestat [52].

**Diabetic retinopathy**

Diabetic retinopathy is the most common cause of blindness in the working age group. Similar to other diabetic microvascular complications, one of the major challenges of preventing diabetic retinopathy is the extremely long latent and silent developmental phase of the disease, where treatment must be initiated before irreversible and devastating consequences occur. Therefore, regular screening of fundus is mandatory in the holistic management of diabetes.

**Glycemia, blood pressure, dyslipidemia, and diabetic retinopathy**

Similar to the beneficial effects on diabetic nephropathy and neuropathy in both the short and long term, intensive glycemic control is also vital in preventing the development and progression of diabetic retinopathy in type 1 diabetic patients as demonstrated by the DCCT and the subsequent 10-year follow-up study [13,53]. This finding further supports the “metabolic memory” phenomenon, as there was a reduced risk of diabetic retinopathy in the intensive treatment group of type 1 diabetic patients, which persisted for 4 years after completion of the DCCT, despite similar HbA1c levels in both the intensive and conventional arms [53].

In the UKPDS, diabetic patients in the intensive treatment group had a 29% risk reduction of retinal photocoagulation compared to those in the conventional group (relative risk 0.71; 95% CI 0.53–0.96; p = 0.0031) [15]. The UKPDS also demonstrates that tight blood pressure control significantly reduces the risk for proliferative retinopathy by 35% [24]. From the results of DIRECT (Diabetic Retinopathy Candesartan Trials), a randomized, double-blinded, placebo-controlled trial in 309 centers worldwide, including 1421 type 1 diabetic patients aged 18–50 years [54], candesartan, an ARB, reduces the incidence of diabetic retinopathy by 35% (hazard ratio 0.65; 95% CI 0.48–0.87; p = 0.0034) but not retinopathy progression (hazard ratio 1.02; 95% CI 0.80–1.31; p = 0.85).

Both high levels of cholesterol and triglycerides have been shown to predict increased rates of diabetic retinopathy. Lipid-lowering agents may protect against retinopathy in diabetic patients by ameliorating oxidative stress and reducing cytokines [55]. The beneficial effects of lipid-lowering agents for the management of diabetic retinopathy are well demonstrated by the FIELD study [40] and the ACCORD Eye Substudy [56]. In the FIELD study, diabetic patients who had poorer glycemic or blood pressure control required laser therapy more frequently. The requirement for first laser therapy was lower in the fenofibrate group compared to the placebo group (hazard ratio 0.69; 95% CI 0.56–0.84; p = 0.0002) [40]. Nonetheless, the mechanism of the beneficial effect of fenofibrate therapy in diabetic retinopathy is unclear, as the effect is not related to the plasma concentrations of lipids. In the ACCORD study, intensive glycemic control and combination treatment with fenofibrate for intensive dyslipidemia therapy, but not intensive blood pressure control, reduced the progression rate of diabetic retinopathy [56]. A subgroup of 2856 type 2 diabetic patients in the ACCORD study was evaluated for the effect of intensive treatment for glycemia and dyslipidemia (fenofibrate 160 mg daily plus simvastatin or placebo plus simvastatin), or systolic blood pressure control treatment (target at <120 vs. <140 mmHg) using two comprehensive, standardized eye examinations conducted by a study ophthalmologist or optometrist, along with fundus photography at baseline and year 4 of follow-up [56]. After 4 years of follow-up, the progression rates of diabetic retinopathy were significantly lower in the intensive glycemic arm (7.3% vs. 10.4%; adjusted odds ratio 0.67; 95% CI 0.51–0.87; p = 0.003). The risk of diabetic retinopathy progression was also reduced in the intensive dyslipidemia group when compared to the conventional group (6.5% vs. 10.2%; adjusted odds ratio 0.60; 95% CI 0.42–0.87; p = 0.006), while there was no difference between intensive blood pressure and conventional groups [56]. In contrast to the results from the DCCT, UKPDS, FIELD, and ACCORD studies, the ADVANCE study failed to show any difference in the incidence of retinopathy (defined as new or worsening retinopathy or visual deterioration) [19].

Again, the STENO-2 study [44] confirms the protective effect of a combination of intensive therapies including blood glucose, blood pressure and lipid control in the prevention of diabetic retinopathy.

**Anti-vascular endothelial growth factor (VEGF) and other treatments in the prevention of diabetic retinopathy**

Since vascular endothelial growth factor (VEGF) has been demonstrated to be extensively involved in the development and progression of diabetic retinopathy, notably diabetic macular edema, VEGF inhibitors may provide an alternative therapeutic approach to retinal laser photocoagulation in the treatment of advanced diabetic retinopathy. A meta-analysis
Good control of glycemia, blood pressure, and lipids↓diabetes-related microvascular complications

Structured care management

Figure 39.2 Evidence from clinical trials supporting structured care management and good control of glycemia, blood pressure, and lipids in primary and secondary prevention of diabetes-related microvascular complications.

of randomized controlled trials in the evaluation of the therapeutic effects of ranibizumab (Lucentis), a recombinant human monoclonal antibody inhibiting VEGF, has shown that ranibizumab is more efficient in reducing the amount of retinal thickening than laser treatment [57]. Recently, ranibizumab was approved for the treatment of diabetic macular edema by the European Medicines Agency and the United States Food and Drug Administration (FDA) [58,59]. Anti-inflammatory agents, including corticosteroid therapy, have also been shown in randomized clinical trials to improve visual acuity in diabetic patients with retinopathy complications [59]. Since oxidative stress is considered to be one of the major contributors to the development of diabetic retinopathy, antioxidants may also play a role in the prevention of diabetic retinopathy, but definitive evidence from clinical trials is still lacking. The role of VEGF inhibitors, anti-inflammatory agents and antioxidants in the primary prevention of diabetic retinopathy in high-risk patients therefore remains to be elucidated.

**Conclusion**

Good control of glycemia, blood pressure, and dyslipidemia are pivotal in the primary and secondary prevention of diabetic microvascular complications as evident from the wealth of data from clinical trials. Structured and comprehensive care is important for management of T2DM, a complex and multifaceted disease (Figure 39.2). Given the growing burden of diabetes and its complications, there is a need for more efficacious treatment strategies to prevent the development and progression of diabetic complications. Future clinical trials should strive to provide conclusive clinical evidence to support or reject the use of novel drugs and treatment regimens to effectively manage diabetes complications.

**References**


SECTION VIII

Management of diabetes: diet, exercise and drugs
Dietary management of diabetes mellitus in Europe and North America

Jim I. Mann1,2 and Alexandra Chisholm1
1Department of Human Nutrition, University of Otago, Dunedin, New Zealand
2Obesity Research Centre, University of Otago, Dunedin, New Zealand

Introduction

Dietary modification is the mainstay of treatment for type 2 diabetes (T2DM). Despite the usually progressive nature of the disease, many newly diagnosed patients with T2DM who comply with dietary advice will show improvement in glycemic control to an extent that will obviate or delay the need for oral agents and insulin. For those who require drug therapy, attention to dietary advice can be expected to further improve glycemic control [1]. In type 1 diabetes (T1DM) dietary advice aims to minimize excessive postprandial glycemia and fluctuations in blood glucose levels. Avoidance of hypoglycemia is a particularly important aim. In all people with diabetes, nutrition therapy is also designed to reduce the risk of long-term complications both by improving glycemic control and by reducing other risk factors (notably dyslipidemia and raised blood pressure levels) for vascular disease. Dietary advice in relation to choices about types of foods for those with T2DM and T1DM is similar, and the major principles resemble those for entire populations at high risk of CHD. There is therefore no need for diabetic patients to be given meals that differ from those eaten by the rest of the family. Dietary recommendations for people with diabetes have been made in many countries. It is reassuring that those of the Diabetes and Nutrition Study Group of the European Association for the Study of Diabetes (DNSG), representing the majority of European countries [2] and the American Diabetes Association (ADA) [3] are generally consistent. These two organizations were amongst the first to issue evidence-based guidelines for the nutritional management of people with diabetes. Any points of difference between the two sets of recommendations are indicated in the text below. Table 40.1 summarizes the basic principles of diabetic dietary recommendations.
Table 40.1 Key aspects of the current recommendations for diabetic diet and lifestyle

<table>
<thead>
<tr>
<th>Dietary energy and body weight</th>
<th>Achieve and/or maintain BMI of 18.5–25 kg m(^{-2}) or if obese, a reduction of at least 5% starting weight. Diet and exercise important.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dietary fat</td>
<td>Saturated plus trans unsaturated fatty acids: &lt; well below 10% total energy. Polysaturated fatty acids: 6–10% total energy. Monounsaturated fatty acids: 10–20% total energy. Total fat: &lt; 35% total energy (if overweight &lt; 30%). Oily fish, soybean and rapeseed oil, nuts and green leafy vegetables to provide n-3 fatty acids. Cholesterol: &lt; 200 g d(^{-1}).</td>
</tr>
<tr>
<td>Carbohydrate</td>
<td>Total carbohydrate: 45–60% total energy, influenced by metabolic characteristics. Vegetables, fruits, legumes, and cereal-derived foods preferred.</td>
</tr>
<tr>
<td>Dietary fiber and glycemic index</td>
<td>Naturally occurring foods rich in dietary fiber are encouraged. Ideally dietary fiber intake should be more than 40 g d(^{-1}) (or 20 g/1000 kcal d(^{-1})), half-soluble (lesser amounts also beneficial). Five servings per day of fiber-rich vegetables and fruit and four or more servings of legumes per week help to provide minimum requirements. Cereal-based foods should be wholegrain and high in fiber. Carbohydrate-rich low-glycemic-index foods are suitable choices, provided other attributes are appropriate.</td>
</tr>
<tr>
<td>Sucrose and other free sugars</td>
<td>If desired and blood glucose levels are satisfactory, free sugars up to 50 g d(^{-1}) may be incorporated into the diet. Total free sugars should not exceed 10% total energy (less for those who are overweight).</td>
</tr>
<tr>
<td>Protein and renal disease</td>
<td>Total protein intake at lower end of normal range (0.8 g kg(^{-1}) d(^{-1})) for type 1 patients with established nephropathy. For all others, protein should provide 10–20% total energy.</td>
</tr>
<tr>
<td>Vitamins, antioxidant nutrients, minerals, and trace elements</td>
<td>Increase foods rich in tocopherols, carotenoids, vitamin C and flavonoids, trace elements and other vitamins. Fruits, vegetables, wholegrains rather than supplements recommended. Restrict salt to less than 6 g d(^{-1}) (less than 2.3 g sodium).</td>
</tr>
<tr>
<td>Alcohol</td>
<td>Up to 10 g for women and 20 g for men per day is acceptable for most people with diabetes who choose to drink alcohol. Special precautions apply to those on insulin or sulfonylureas, those who are overweight and those with hypertriglyceridemia.</td>
</tr>
<tr>
<td>Special “diabetic” foods, or functional foods and supplements</td>
<td>Nonalcoholic beverages sweetened with non-nutritive sweeteners are useful. Other special foods not encouraged. No particular merit of fructose and other “special” nutritive sweeteners over sucrose.</td>
</tr>
<tr>
<td>Families</td>
<td>Most recommendations suitable for whole family.</td>
</tr>
</tbody>
</table>

**Total energy intake**

The majority of people with T2DM are overweight and usually have other features of the metabolic syndrome (raised levels of insulin, raised triglyceride and low high-density lipoprotein (HDL) levels, hypertension, high levels of uric acid, and elevated levels of plasminogen activator inhibitor 1) in addition to hyperglycemia. Even modest weight reduction is associated with a reduction in insulin resistance, and hepatic glucose production, and perhaps an improvement in islet β-cell function. A reduction in blood glucose may be seen with reduced energy intake even before weight loss occurs. In addition to the resultant improvement in overall glycemic control, as measured by HbA1c, many of the other abnormalities associated with the metabolic syndrome, notably dyslipidemia and raised blood pressure, also improve with weight loss. The improvement in metabolic parameters associated with weight loss is seen not only in the early stages of the disease but has also been demonstrated in patients in whom glycemic control was considered to have been optimized on standard treatment regimens involving insulin and or oral hypoglycemic agents [1].

It is generally accepted that total energy (TE) should be reduced so that BMI moves toward the desirable range (20–25 kg m\(^{-2}\)). Some Asian populations which are well represented in Europe and North America have a relatively high proportion of visceral fat and central adiposity and an even higher risk of cardiovascular disease than European populations with diabetes. For them a slightly lower BMI is generally regarded as optimal. However, despite advice and support many people with T2DM are unable to achieve this target. It is important therefore to set realistic goals for weight reduction and to provide intermediate as well as long-term targets.
A sustained weight loss of 1–2 kg per month with the aim of achieving a long-term weight loss of 5–7% of starting weight is a reasonable goal for most patients. This can usually be achieved by a reduction of about 500 kcal below that required for weight maintenance. Weight loss may lead to greater improvements in cardiac risk factors in individuals with a high waist circumference but all patients with excessive adiposity, regardless of distribution, should be advised to lose weight. Regular physical activity is generally considered to enhance weight loss and to be particularly important in weight maintenance when targets have been achieved. Behavior modification techniques are an important component of weight loss programs.

A wide range of “diets” of varying macronutrient composition (e.g., low carbohydrate, high protein, high carbohydrate-high fiber, high fat) and dietary patterns (e.g., Mediterranean) have been promoted as being the most appropriate options for achieving weight loss in people with diabetes. Appropriate distribution of macronutrients is considered in the sections which follow. However, it should be noted that there is no clinical trial evidence that in the medium- to long-term, any one dietary approach has advantages over the others with regard to either weight loss or clinical outcomes [4,5]. It is particularly noteworthy that diets very low in carbohydrate may be associated with increases in LDL-cholesterol and generally involve the elimination of foods that are important sources of fiber and essential micronutrients [5]. Carbohydrate intakes below about 130 g d\(^{-1}\) are not advised.

The most appropriate approach to achieving weight reduction depends upon the requirements of individual patients. For some it may be sufficient to advise the restriction of energy-dense foods and provide advice with regard to appropriate food types, as well as to provide some general guidance concerning increased physical activity. For those unable to lose weight, a more prescriptive approach based on calorie-counting as well as an exercise prescription may be helpful, especially when this approach is used in conjunction with behavioral modification techniques.

The role of liquid or solid packaged meal replacements intended to be used instead of one or two regular meals per day can facilitate substantial energy reduction. However their place, as well as the role of very low calorie diets (VLCDs), which provide 800 or fewer calories daily in the management of obesity in diabetes, remains to be established. There is no doubt that they can be associated with rapid and substantial weight loss with impressive improvement in glycemic control. The long-term benefits are less clear since weight regain frequently occurs when normal meals are reintroduced. There is some evidence that despite greater initial weight loss, when VLCDs are used for induction of weight loss, a conventional program is associated with greater sustained reduction in weight and improvement in several clinical and metabolic measures. When VLCDs are used, such as in the management of severe obesity, this should be under the supervision of experienced physicians and dietitians or nutritionists since major alterations in hypoglycemic therapy may be needed and care taken to ensure an adequate intake of essential nutrients. Long-term dietary support, including behavioral modification techniques are essential [6]. Use of existing weight loss drugs has little or no place in the management of overweight and obesity in diabetes. Appropriately selected patients typically have impressive reductions in body weight and dramatic improvements in glycemic control and some cardiovascular risk factors following bariatric surgery.

For those with T2DM who are not overweight, and for the majority of those with T1DM, self-selected energy intakes are appropriate. It is important to ensure that energy intake is sufficient to achieve growth and development in childhood and adolescence. The prescription of insulin in excess of requirements may lead to weight gain. Overweight patients with T1DM can often appreciably improve glycemic control and at the same time reduce insulin requirements if they are able to lose weight. With increasing rates of obesity in North America and most European countries it is hardly surprising that an increasing number of people with T1DM are also presenting with overweight and obesity and associated insulin resistance. It is important that this situation be recognized and managed with advice to lose weight.

### Quantity and nature of dietary fat

The recommendations on dietary fat intake are based on epidemiologic studies in nondiabetic subjects and those with diabetes and controlled dietary intervention studies involving glycemic control and cardiovascular disease risk factors as surrogate endpoints. Most risk factors for cardiovascular disease first identified in nondiabetic subjects also operate in people with diabetes. The striking relationship between saturated and trans unsaturated fatty acids and coronary heart disease [7] and the adverse effect of these fatty acids on the atherogenic low-density lipoproteins and lipoprotein [6,8] and insulin sensitivity provide convincing evidence for the recommendations to appreciably reduce their intakes. Saturated fatty acids may also induce a detrimental postprandial lipid profile compared with monounsaturated fat in nondiabetic and diabetic subjects [9] and have the potential to increase postprandial insulinemia in obese subjects with T2DM [10]. Exchange of saturated fatty acids with unsaturated fatty acids improves insulin sensitivity [8]. The European recommendations suggest a reduction in saturated and trans unsaturated fatty acids to below 10% TE, or to less than 8% if LDL-cholesterol is elevated. The ADA recommendations suggest that all people with diabetes should reduce their intake of saturated fat to below 7% and that intake of trans fat should be minimized. This represents appreciable reductions, given that intakes are currently up to twice this level in North America and Europe [11]. Based on research in nondiabetic individuals, it appears that myristic and palmitic acids have the greatest adverse effect on LDL-cholesterol and stearic acid has little or no effect. Lauric acid has an intermediate effect [12].
Partial substitution of \(n\)-6 polyunsaturated fatty acids (chiefly linoleic acid) for saturated fatty acids may help to reduce LDL-cholesterol and has been associated with reduced risk of coronary heart disease events [13] but in large quantities (i.e., more than 10% of TE) may reduce HDL and increase lipid peroxidation. \(n\)-3 Polyunsaturated fatty acids (eicosapentaenoic and docosahexaenoic acids), chiefly derived from oily fish, can help to reduce triglycerides and VLDL, as well as reduce the risk of thrombosis as a result of reduced platelet aggregation. However, these fatty acids have a variable effect on LDL and the most appropriate ratio of \(n\)-3 to \(n\)-6 unsaturated fatty acids has yet to be established. Total intake of polyunsaturated fatty acids is very low in most European countries and North America and some increase may be appropriate, though total intake should not exceed 10% of TE. Fish oil supplements are not advised for routine use although increased intake of fish will facilitate the reduction of saturated fat as well as increase the intake of naturally occurring \(n\)-3 fatty acids, especially if oily fish is eaten regularly. Both the ADA and European recommendations include advice to consume 2–3 servings of fish each week and the European recommendations also encourage consumption of plant sources of \(n\)-3 fatty acids (e.g. rapeseed oil, soyabean oil, nuts, and some green leafy vegetables).

Monounsaturated fatty acids with a \(cis\) configuration (principally oleic acid from olive oil and the rapeseed oil derivative, canola) on the other hand, appear not only to facilitate the reduction of saturated fatty acids (SAFA), but may be associated with an overall more favorable lipid profile when compared with carbohydrate-rich food used as replacement for saturated fats [14]. There is also some evidence that blood pressure levels may be reduced and peripheral insulin sensitivity improved on diets relatively high in monounsaturated fatty acids. A fairly wide range of intakes (10–20% TE) is considered acceptable resulting in an appreciably wider acceptable range of intakes for total fat than was permitted in previous recommendations. However, given the contribution of fat regardless of its nature to the overall energy density of the diet, it is recommended that total fat intake should not exceed 35% TE or 30% TE for those who are overweight. The degree of adiposity, metabolic characteristics and personal and cultural preferences all contribute to the decision as to the optimal quantity and nature of total fat. It has been suggested that monounsaturated fatty acids may be more appropriate sources of replacement energy for saturated fatty acids than carbohydrates, given that high intakes of some carbohydrate-rich foods may result in increased levels of triglycerides and reduced HDL. Such advice needs to take into account the high energy density of all fats and the potential of fiber-rich carbohydrate foods to promote satiety. This is further discussed in a later section of the chapter.

Substantial reduction of saturated fat will almost invariably be associated with a reduction of dietary cholesterol, to 200 mg or less per day, the intake suggested in the ADA recommendations. There is no good evidence to suggest that a further reduction of dietary cholesterol confers additional benefit when intake of saturated fat is already low.

**Protein**

Protein intake in Europe and North America ranges between 10% and 20% TE which corresponds to 1.3–2.0 g kg\(^{-1}\) body weight. This exceeds requirements. In T1DM albumin excretion rate increases with protein intake beyond this level especially in the presence of poor glycemic control and/or hypertension. The recommended range of intake in patients with no evidence of nephropathy is 10–20% TE in Europe and 15–20% TE in the United States. In patients with T1DM and established nephropathy there is evidence that limiting protein intake to the lower end of the recommended range of intakes (0.8 g kg\(^{-1}\) body weight per day) improves outcome measured as reaching end-stage renal disease or death [15]. Thus in this group of patients, this level of restriction is recommended. For patients with T1DM and incipient nephropathy (microalbuminuria) and those with T2DM and established or incipient nephropathy, there is insufficient evidence to make firm recommendations regarding protein restriction. There is also insufficient evidence to make recommendations about preferred protein type.

Recently published evidence and the availability of several popular diet books have led to the suggestion that high protein diets (20–30% TE) which may also be high in fat and relatively low in carbohydrate are particularly helpful in facilitating weight reduction and improving the metabolic derangements associated with insulin resistance, notably high triglyceride and low HDL-cholesterol levels [16]. The only long-term study in people with T2DM which attempted to implement a diet with dietary protein in this range found that the majority of patients did not comply with the high protein advice over an 18-month period [17]. In that study a modest improvement in glycemic control occurred in both the high protein and the relatively high carbohydrate control diet and was associated with similar modest weight loss which occurred on both diets. While it seems reasonable, especially for overweight or obese patients with T2DM, to suggest protein intakes at the upper end of the recommended range (i.e., 20%TE), in order to facilitate reduction of rapidly digested carbohydrate (see section that follows) higher intakes are not currently advised. Long-term effects of high protein diets in diabetes are not known.

**Carbohydrate, dietary fiber, and glycemic index**

Traditionally in many affluent societies, low carbohydrate diets (40% or less TE) were recommended for people with diabetes. This was inevitably associated with an increase in dietary fat, especially saturated fat. With increasing evidence for the
deleterious effects of saturated fatty acids and the potential for dietary fiber, especially soluble gel-forming fiber, to improve glycemic control and some cardiovascular risk factors, there was a trend towards recommending much higher intakes of carbohydrate, up to 60% TE [18]. What was frequently not appreciated was that the benefits of a high carbohydrate, high fiber diet had been demonstrated in studies involving very high intakes of fiber-rich carbohydrate (40g dietary fiber or more per day) typically derived from intact fruit and vegetables, legumes, and lightly processed wholegrain cereals [19]. This advice regarding the acceptability of high carbohydrate intakes was frequently translated, by patients and sometimes health professionals, into diets high in rapidly digested starchy foods such as rice, potato, and “wholemeal” bread. There was no evidence that such foods conferred benefit in terms of glycemic control. Indeed it is unsurprising that a meta-analysis comparing the effect of high carbohydrate diets (49–60% carbohydrate, 20–32% fat, 7–13% MUFA) with diets higher in monounsaturated fat and lower in carbohydrate (36–40% carbohydrate, 37–50% fat, 22–33% monounsaturated fatty acids (MUFAs) found that triglyceride levels were appreciably lower on the high MUFAs diets. Fasting glucose levels were slightly lower and HDL-cholesterol levels slightly higher on the high MUFAs diets than the high carbohydrate diets which were typically rich in starchy foods [20]. Thus the current European recommendations specify a wide range of acceptable intakes for total carbohydrate (45–60% TE) which is based on the limits for total fat and protein intakes. They specify that metabolic characteristics suggest the most appropriate intake within this range. For those with marked insulin resistance and appreciably raised triglyceride levels, intakes towards the lower end of the range are preferred.

Source of dietary carbohydrate is particularly important. Both European and US recommendations encourage a dietary pattern that includes carbohydrate from a range of fruits, vegetables, and wholegrains with particular emphasis on legumes which are rich in soluble fiber. For those who choose an intake at the upper end of the recommended range, it is particularly important to emphasize foods rich in dietary fiber. Dietary fiber is emphasized in both sets of recommendations with the European set suggesting 20 g/1000 kcal or more if possible, about half of which should be soluble. Daily consumption of at least five servings of fiber-rich vegetables or fruit, and at least four servings per week of legumes help to provide minimum requirements and cereal-based foods should, whenever possible, be wholegrain and high in fiber.

Low carbohydrate diets in which carbohydrate is restricted to below 130 g d⁻¹ are not recommended, both because such diets are likely to be unacceptably high in fat and because carbohydrate-containing foods are important sources of vitamins and minerals.

Postprandial glycemia is influenced by amount and type of carbohydrate. Carbohydrate counting is a helpful approach for patients with T1DM and those with T2DM who are treated with insulin, to determine insulin dose required to cover a carbohydrate-containing meal or snack [21]. However, nature of carbohydrate is also a determinant of glucose levels after a meal. The glycemic index (GI), (defined as the incremental blood glucose area following ingestion of 25–50 g available carbohydrates expressed as a percentage of the corresponding area following ingestion of carbohydrate from a reference food, glucose, or white bread) of carbohydrate-containing foods provides an indication of the extent of expected glucose rise following their consumption [22]. This in turn depends upon the speed, extent, and site of digestion and absorption. Examples of the GI values of some carbohydrate-containing foods is shown in Table 40.2. Rapidly digested starchy foods and some sugars tend to have high glycemic indices though sucrose, high fructose corn syrups (HFCS), and foods rich in both do not, because fructose (sucrose contains 50% fructose and HFCS, 55% fructose) is very rapidly metabolized and does not influence blood glucose levels. While the GI of carbohydrate-containing foods does provide some indication as to their glycemic potential, and there is some evidence that emphasis on low GI foods can improve overall glycemic control, [23] the index does have a number of important limitations [24]. These include the appreciable inter-and intra-individual variation and the fact that the index provides no indication as to the overall impact of the food. For example, energy-dense foods high in fat and sugar such as ice-cream tend to have a low glycemic index. However, when considering similar products such as bread, the choice of a low GI brand may be helpful, especially when the choice is made also taking into account other attributes such as fiber and sodium content. As a general rule, the index should not be used for foods such as ice-cream which are also high in fat. Sucrose and other free sugars may be incorporated into the diet provided contribution to TE is taken into account and intake does not exceed 50 g d⁻¹.

Dietary fiber supplements are not generally recommended.

Distribution of energy from macronutrients

The major dietary restriction is that of saturated and trans unsaturated fatty acids to no more than 10% of TE, or 7% if LDL-cholesterol is raised according to the European recommendations. The ADA recommendations suggest that all people with diabetes should derive less than 7% TE for saturated and trans fatty acids should be restricted. If energy restriction is not required (i.e., if BMI is in the satisfactory range) the lost energy may be replaced by either an increase in cis monounsaturated fatty acids from vegetable sources (e.g. olive oil, canola) or carbohydrates. Both these nutrients have a wide acceptable range of intakes: cis monounsaturated fatty acid, 10–20% of TE, total carbohydrates, 45–60% of TE. This involves a restriction of fatty meat and meat products as well as high-fat dairy products, confectionery, cakes, and high-fat manufactured foods, which
are replaced by monounsaturated vegetable oils and spreads and fiber-rich or carbohydrates. It is with regard to nature of carbohydrate that there has been an apparent discrepancy between North American and European recommendations. Early versions of the ADA recommendations, while acknowledging the health benefits of “foods containing carbohydrate from wholegrains, fruit, vegetables and low-fat milk,” suggested that the total amount of carbohydrate was more important than source and type, that sucrose did not increase glycemia to a greater extent than isocaloric amounts of starch and that restriction by people with diabetes was therefore unnecessary. European recommendations are much more directional and stated clearly that vegetables, legumes, fruits, and cereal-derived foods were the preferred carbohydrate choices and that whenever possible, they should have a low glycemic index and be high in dietary fiber. The more recent ADA guidelines are now more specific in terms of glycemic index and dietary fiber so the discrepancies between the two sets of recommendations are much less marked. In most European countries and North America, intake of n-3 and n-6 polyunsaturated fatty acids is relatively low and some increase up to a maximum of 10% TE may be appropriate. It is important to emphasize that the overall distribution of macronutrients is no different from that widely advised for the population as a whole in countries with high coronary heart disease (CHD) risk. For those who are overweight it is appropriate to recommend only partial replacement of energy lost by reduced intake of saturated and trans unsaturated fatty acids. A number of overweight patients have opted to try very low carbohydrate, high fat or high protein dietary approaches to weight loss (e.g. the Atkins Diet, Zone diet). While such diets often achieve short-term weight loss and associated benefits, including dose reduction of hypoglycemic drug therapy, extreme dietary practices should be discouraged at least until such a time as they have been subjected to further appropriate evaluation. At present there is no evidence that in the long term they are more efficacious than established dietary approaches, and there is at least a theoretical basis for believing that there may be some deleterious long-term effects. Protein intakes at the upper end of the recommended range (20% TE) and carbohydrate at the lower end (45% TE) are acceptable and may be particularly appropriate for overweight patients and those with features of the metabolic syndrome.

It should be clear from this discussion that the acceptable range of intakes of macronutrients is compatible with the wide range of dietary patterns traditionally consumed by people throughout Europe and North America. The Mediterranean diet typically eaten in southern Europe is very different from that consumed in the Nordic countries, yet both these dietary patterns, as well as the “typical” British diet consisting of meat and three vegetables at main meals, can all be prepared so that macronutrient distribution conforms with the recommended ranges. This applies also to the wide variety of Asian cuisines which are now popular in Europe and North America. However, it should be noted that while using traditional methods

Table 40.2  Mean glycemic index of a selection of foods

<table>
<thead>
<tr>
<th>Breads</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>78</td>
</tr>
<tr>
<td>Wholemeal</td>
<td>80</td>
</tr>
<tr>
<td>Mixed grain</td>
<td>58</td>
</tr>
<tr>
<td>Pumpernickel</td>
<td>56</td>
</tr>
<tr>
<td>Wholemeal rye</td>
<td>66</td>
</tr>
<tr>
<td>Rye crispbread</td>
<td>64</td>
</tr>
<tr>
<td>Cereal grains</td>
<td></td>
</tr>
<tr>
<td>Bulgur</td>
<td>48</td>
</tr>
<tr>
<td>Barley</td>
<td>28</td>
</tr>
<tr>
<td>Buckwheat</td>
<td>54</td>
</tr>
<tr>
<td>Quinoa</td>
<td>53</td>
</tr>
<tr>
<td>Rice</td>
<td></td>
</tr>
<tr>
<td>Jasmine, cooked in rice cooker</td>
<td>68</td>
</tr>
<tr>
<td>Basmati, cooked in rice cooker</td>
<td>57</td>
</tr>
<tr>
<td>Doongara, cooked in rice cooker</td>
<td>55</td>
</tr>
<tr>
<td>Brown, cooked in rice cooker</td>
<td>65</td>
</tr>
<tr>
<td>Parboiled, cooked in rice cooker</td>
<td>57</td>
</tr>
<tr>
<td>Pasta</td>
<td></td>
</tr>
<tr>
<td>Fettucini</td>
<td>32</td>
</tr>
<tr>
<td>Spaghetti, white, boiled</td>
<td>46</td>
</tr>
<tr>
<td>Spaghetti, wholemeal, boiled</td>
<td>42</td>
</tr>
<tr>
<td>Macaroni, plain, boiled</td>
<td>47</td>
</tr>
<tr>
<td>Breakfast cereals</td>
<td></td>
</tr>
<tr>
<td>Oatmeal</td>
<td>55</td>
</tr>
<tr>
<td>Special K™ (Kellogg’s)</td>
<td>69</td>
</tr>
<tr>
<td>All-Bran™</td>
<td>55</td>
</tr>
<tr>
<td>Cornflakes™</td>
<td>93</td>
</tr>
<tr>
<td>Coco Pops™</td>
<td>77</td>
</tr>
<tr>
<td>Legumes</td>
<td></td>
</tr>
<tr>
<td>Baked beans, canned</td>
<td>60</td>
</tr>
<tr>
<td>Chick peas, dried boiled</td>
<td>36</td>
</tr>
<tr>
<td>Chick peas, canned in brine</td>
<td>42</td>
</tr>
<tr>
<td>Green peas, frozen, boiled</td>
<td>51</td>
</tr>
<tr>
<td>Kidney beans, dried boiled</td>
<td>29</td>
</tr>
<tr>
<td>Red lentils, dried boiled</td>
<td>26</td>
</tr>
<tr>
<td>Haricot and navy beans, dried boiled</td>
<td>31</td>
</tr>
<tr>
<td>Butter beans, dried cooked</td>
<td>31</td>
</tr>
<tr>
<td>Root vegetables</td>
<td></td>
</tr>
<tr>
<td>White potatoes, boiled</td>
<td>70</td>
</tr>
<tr>
<td>Mashed potatoes</td>
<td>74</td>
</tr>
<tr>
<td>Sweet potatoes, kumara (New Zealand)</td>
<td>61</td>
</tr>
<tr>
<td>Yam, boiled (New Zealand)</td>
<td>37</td>
</tr>
<tr>
<td>Fruit</td>
<td></td>
</tr>
<tr>
<td>Apple</td>
<td>38</td>
</tr>
<tr>
<td>Banana</td>
<td>62</td>
</tr>
<tr>
<td>Orange</td>
<td>40</td>
</tr>
<tr>
<td>Kiwifruit</td>
<td>47</td>
</tr>
<tr>
<td>Pear</td>
<td>38</td>
</tr>
<tr>
<td>Grapes</td>
<td>59</td>
</tr>
</tbody>
</table>

of preparation many dishes are indeed suitable for people with diabetes, but when traditional recipes have been adapted for Western consumption they may have increased amounts of fats, sugars, and rapidly digested starches.

**Antioxidant nutrients, vitamins, and minerals**

A whole range of micronutrients have at one time or another been the focus of attention in diabetes. Dietary sodium restriction can produce substantial reduction in blood pressure in mildly hypertensive patients with T2DM [25] and to enhance the blood pressure lowering effects of other dietary manipulations (low-fat dairy products, fruits, and vegetables) in non-diabetic individuals [26]. So all people with diabetes are advised to restrict salt intake to under 6 g d$^{-1}$. Those with elevated blood pressure may be recommended to further reduce their intake.

There is a substantial body of evidence from prospective studies in nondiabetic individuals which suggests that a range of antioxidant nutrients and good food sources of these (fruits, vegetables, nuts, berries) are protective against cardiovascular disease [27,28]. People with diabetes may be in a state of increased oxidative stress. There is some evidence from short-term studies that consumption of foods rich in these nutrients and supplementation with a range of synthetic micronutrients that some markers of oxidative stress can be favorably influenced by substantial intakes [29]. However, there is no convincing evidence from studies with clinical endpoints and even some evidence for potential harm with high doses of antioxidant supplements. Thus it seems appropriate to recommend foods rich in dietary antioxidants, but supplements are not recommended.

There has also been interest in possible roles for chromium, potassium, magnesium and zinc, deficiencies of which may aggravate carbohydrate intolerance. Deficiencies, especially of magnesium, have been detected in poorly controlled type 1 patients. There is no convincing evidence for supplementation with these or any other minerals, vitamins or trace elements unless deficiency has been established.

**Sweeteners, supplements and functional foods**

A range of nonnutritive sweeteners is available for use in drinks, cooking, and manufactured foods. Aspartame is generally included in this category although it is a dipeptide, because it is so sweet that only very small quantities are required. These sweeteners provide a useful means of reducing energy intake.

Several nutritive sweeteners, notably fructose and sorbitol have also been used extensively for home use and manufactured foods in so-called “diabetic foods” as an alternative to sucrose. They are not associated with postprandial glycemia but have the same (fructose) or half (sorbitol) the energy content as sucrose and may be associated with untoward effects. For example in large quantities fructose may cause diarrhea and hypertriglyceridemia. However, while they have no proven advantage over sucrose, moderate intakes (up to a maximum of 50 g d$^{-1}$) do not appear to have adverse gastrointestinal or metabolic effects.

Many functional foods and supplements are promoted as beneficial for the management of people with diabetes or reducing the risk of complications. Some have been shown to have potentially relevant functional effects but have not been tested in long-term clinical trials. Most are expensive and many are relatively high in energy. While sugar-free and low-calorie products may be useful for many people with diabetes, functional foods and supplements are not generally recommended in “official” recommendations for people with diabetes. It is generally accepted that the principal benefits of nutritional approaches to the treatment (and prevention) of diabetes are derived from the appropriate intakes of usual foods.

**Alcohol**

The same caution regarding alcohol intake that applies to the general population applies to those with diabetes. Alcohol consumption should be eliminated as far as possible in those with a history of pancreatitis, hypertriglyceridemia, those who are overweight, and those with hypertension. When alcohol is taken by those on insulin or tablets, it should be taken with meals because of its potentially profound hypoglycemic effect, which can be prolonged or delayed. There is evidence, principally in people who do not have diabetes, that a modest intake of alcohol may reduce cardiovascular risk because of a beneficial effect on HDL-cholesterol, reduced coagulability, and reduced lipid oxidation [30,31]. However, this evidence is not considered sufficient to justify a recommendation to drink alcohol if individuals are not already doing so. For those who choose to drink alcoholic beverages and do not have the complications listed above, a quantity equivalent to 10 g d$^{-1}$ for women and 20 g d$^{-1}$ for men is acceptable.

**Implementation of diet therapy**

The dietary prescription for people with diabetes often involves a major change in lifestyle and physicians are rarely competent to be the sole providers of appropriate advice, though their support in emphasizing the importance of nutrition therapy is critical in encouraging compliance with the advice given by dietitians and appropriately trained nutritionists. Behavioral techniques are almost always needed in addition to the practical advice regarding food purchase and preparation required to translate nutrition principles into practice. Implementation of current nutrition therapy should be easier than in the past since the principles of diet composition are those recommended for
Table 40.3 Guidelines for optimal food choices

<table>
<thead>
<tr>
<th>Eat regularly (servings to aim for/serving size)</th>
<th>Eat in moderate amounts</th>
<th>More often</th>
<th>Occasionally</th>
<th>Seldom or avoid</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pulses &amp; legumes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2-3 per week/1 cup cooked)</td>
<td></td>
<td>Peanuts (unsalted) (see under nuts below)</td>
<td>Refried beans cooked in a little suitable MUFA or PUFA oil</td>
<td>Deep-fried legumes</td>
</tr>
<tr>
<td>Beans (e.g. haricot, red kidney, soya, baked, black-eyed)</td>
<td>Falafel baked or &quot;dry fried&quot;</td>
<td>Low-fat refried beans</td>
<td>Falafel baked or fried in a little suitable MUFA or PUFA oil</td>
<td></td>
</tr>
<tr>
<td>Lentils</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chick-peas, split peas</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>**Fruit—**with skin where possible</td>
<td>Fruit stewed without sugar or canned in water or fruit juice (as part of fruit recommendation) (1/2 to 2/3 cup)</td>
<td></td>
<td>Dried fruit</td>
<td>Fruit canned in heavy syrup</td>
</tr>
<tr>
<td>(2-4 per day/1 piece raw)</td>
<td></td>
<td></td>
<td>Fruit juice (diluted)</td>
<td>Fruit leather</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Crystallized fruit</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Fruit juice (undiluted)</td>
</tr>
<tr>
<td><strong>Vegetables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colored vegetables</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3-5 per day/1/2 to 1 cup)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fresh or frozen, cooked (e.g. carrots, cauliflower, cabbage, spinach, green beans, peas, corn, beetroot, zucchini)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salad or raw vegetables (e.g. tomato, lettuce, cucumber, radish, celery, capsicum)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Breads and cereals</strong></td>
<td>White bread, white flour</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At least 1/2-1/2 wholegrain</td>
<td>Unsweetened breakfast cereals (e.g. Special K™, Cornflakes™)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breads &amp; rolls (1 slice/1 small roll)</td>
<td>Sugar-coated cereals (e.g. Sugar Puffs™, Frosties™, Coco Pops™)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cereals (1/3-1 cup)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oats, barley</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wholegrain breads and rolls</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pumpernickel bread</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low-fat wholemeal crispbreads</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wholegrain cereals</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unsweetened breakfast cereals (e.g. All Bran™, Weetbix™, Weetabix™, wholegrain porridge, Bircher muesli—made with rolled oats low-fat yoghurt &amp; low-fat milk, grated apple (a little dried fruit &amp; nuts—optional), oat-based unsweetened muesli)</td>
<td>Jasmine rice</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brown rice, wholegrain rice</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Doongara, Basmati rice</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wholemeal pasta, wholemeal flour, rye flour</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oatcakes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Category</td>
<td>Examples</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>--------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| **Meat, poultry, and fish** | **Meat & poultry (3–4 per week/50–120 g cooked weight)**  
Lean meat or poultry, cooked by low-fat cooking methods (baked, microwaved, grilled)  
Shellfish or other seafood cooked without fat (poached, baked, microwaved, grilled, steamed)  
Lean meat or poultry, roasted or cooked in a little MUFA or PUFA oil  
Meat fat, poultry fat or skin, fried meat, duck, goose, meat pies, fatty mince, hot dogs, sausages  
Luncheon meat, salami, canned corned beef, fatty bacon, crackling  
Deep-fried fish, or seafood  
Fish & meat paté |
| **Fish** (1–2 per week/100–200 g cooked weight) | **White fish**  
Some oily fish (50–75 g) (e.g. salmon, tuna, sardines tinned in water)  
**Eggs** (1 egg or egg white)  
**Egg whites**  
**Dairy products—low fat (LF)** (2–3 per day/100–150 mL)  
**LF/low-sugar yoghurt**  
**Skimmed milk**  
**Cottage cheese (40 g)**  
**Curd cheese (serving size 20 g)**  
**Low-fat quark**  
**Homemade fruit smoothies made with skimmed milk**  
**Semi-hard cheese, mozzarella, gouda, edam, brie, feta, (serving size 30 g)**  
**Plain ice-cream, full-fat sweetened yoghurt, whole milk, flavored milk**  
**Reduced fat cream**  
**Hard cheese e.g. cheddar, Tasty, blue cheese (serving size 20 g)**  
**Double-cream brie and camembert**  
**Lite cream cheese**  
**Lite sour cream**  
**Butter 1/2 – 1 tsp**  
**Peanut butter 1 – 1 1/2 tsp** (no added sugar or salt)  
**1 tablespoon pesto**  
**1 tablespoon olives**  
**Margarine/butter mixture 1 tsp**  
**Soft margarine 1 tsp**  
**Condensed milk**  
**Sour cream**  
**Milk-shakes with sugar syrup flavoring**  
**Lard, suet, dripping**  
**Oils: coconut, palm**  
**Partially hydrogenated oils hydrogenated fats (e.g. hard margarines)**  
| **Fats and oils (1–2 tsp)** | Use in place of spreads high in animal fats: cis monounsaturated oils (e.g. olive or canola)  
cis polyunsaturated oils (e.g. sunflower, soya, grapeseed)  
Soft margarine 1 tsp  
Margarine/butter mixture 1 tsp  
| **Nuts and seeds** | **Nuts**  
Use in place of foods high in animal fats (3–5 per week/20–30 g), (e.g. almonds, pistachios, Brazil nuts, hazelnuts, cashews, macadamias, peanuts, pine nuts, walnuts, pecans (nuts should be unsalted, and not roasted in other oils or fats))  
Use in place of foods high in animal fats  
Nut butters (no added sugar or salt)  
1 – 1 1/2 tsp  
| **Seeds** | 1 tbsp (e.g. sunflower, sesame, pumpkin, flaxseed)  
Avocado 1 tbsp  
| **Continued overleaf** |
### Table 40.3 (continued)

<table>
<thead>
<tr>
<th>Eat regularly (servings to aim for/serving size)</th>
<th>More often</th>
<th>Occasionally</th>
<th>Seldom or avoid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Made-up dishes</td>
<td>“Cream” soups made with skim milk&lt;br&gt;Sauces made with skim milk and wholemeal flour</td>
<td>Filo pastry a little suitable MUFA or PUFA oil&lt;br&gt;(between several sheets not each sheet)&lt;br&gt;Pastry made with wholemeal flour and MUFA or PUFA oils or soft margarines&lt;br&gt;Low-fat, low-sugar, high-fiber desserts, cakes, or biscuits made from special recipes, with wholemeal flour and soft MUFA or PUFA margarine</td>
<td>Cream soups, sauces with cream&lt;br&gt;Flaky or short pastry&lt;br&gt;Meat or pork pies&lt;br&gt;Sausage rolls&lt;br&gt;Sweet pastries&lt;br&gt;High-fat, high-sugar desserts</td>
</tr>
<tr>
<td>Home-made vegetable or clear soups</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Takeaways</td>
<td>Wholemeal, wholegrain rolls or wraps or bread with very lean meat or poultry or vegetarian fillings and assorted salad fillings</td>
<td>Pasta (small serving with tomatoes, chicken, lean meat, seafood or pesto)&lt;br&gt;Steamed rice or noodles, plus chicken, fish or beef with vegetables&lt;br&gt;Chicken or meat cooked in a sauce without butter or cream&lt;br&gt;Burger grilled meat patty and salad&lt;br&gt;Grilled chicken or meat with salad&lt;br&gt;Oven-baked chips, wedges&lt;br&gt;Pizza (small) with tomato, vegetable, lean meat, chicken or fish, very little cheese</td>
<td>Deep-fried takeaways&lt;br&gt;Pasta with cream sauce&lt;br&gt;Cream or butter sauces&lt;br&gt;Thin crisps&lt;br&gt;Salami&lt;br&gt;Extra cheese</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spreads and preserves</td>
<td>Marmite™, Promite™</td>
<td>Jam, marmalade</td>
<td>Sugar, glucose, honey, syrup, treacle, Sweets, chocolates, fudges, butterscotch&lt;br&gt;Fruit mincemeat&lt;br&gt;Lemon curd&lt;br&gt;Sweet sauces &amp; toppings</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drinks</td>
<td>Water, tea, coffee, tomato juice, sugar-free drinks (e.g. soda water, sugar-free squash)</td>
<td>Spirits with diet mixers, dry wine, &amp; sherry&lt;br&gt;French champagne&lt;br&gt;Low-calorie beer</td>
<td>Sweet wines, liqueurs&lt;br&gt;Sweet sherry, sweetened fruit drink&lt;br&gt;Sweetened fizzy drinks&lt;br&gt;High-fat drinks with alcohol &amp; cream&lt;br&gt;Pre-mixed drinks&lt;br&gt;Regular beers</td>
</tr>
</tbody>
</table>

Source: Dr. Alex Chisholm. Reproduced with permission of Dr. Alex Chisholm.
healthy eating and the foods required by those with diabetes should not differ appreciably from those appropriate for the entire family.

Most individuals with T1DM will be using rapid-acting insulin as a component of a basal-bolus insulin regime or via an insulin pump. For them, the total carbohydrate content of meals and snacks, typically estimated by carbohydrate counting or experience-based estimation, together with pre-meal blood glucose level will guide the dose of the insulin bolus [21, 32]. This method enables a more flexible approach than the earlier system of prescribing a fixed number of carbohydrate exchanges to match prescribed insulin dose. For those on fixed insulin regimens total carbohydrate will need to be kept fairly consistent with regard to time and amount on a day-to-day basis in order to achieve satisfactory glycemic control.

Some people with T2DM on intensive insulin regimes will also require advice with regard to timing and amount of total carbohydrate. However, for the majority, emphasis will be on dietary modifications that ensure appropriate energy intakes, carbohydrate. However, for the majority, emphasis will be on dietary modifications that ensure appropriate energy intakes, reductions of saturated and trans fatty acids and sodium and adequate intake of other nutrients. While individualized specialist dietary advice from dietitians, taking into account individual preferences and metabolic characteristics, is obviously desirable for all people with diabetes, this will not be universally available. Many people with T2DM may be able to cope adequately with generalized advice about nutritional and dietary principles (Table 40.3) provided they receive the support of other available health professionals (e.g. specialist diabetes nurses or trained community workers) and family members. Regular follow-up by nonspecialist nurses can also encourage dietary compliance and has been shown to support maintenance of weight loss over a prolonged period [33].

While the role of physical activity in the prevention and treatment of diabetes has received considerable attention [34] the early ACTID randomized control has shown that the addition of an activity intervention conferred no additional benefit to that observed in the group that received only intensive dietary advice [35].

References


CHAPTER 41

Exercise

Henrik Galbo1 and Erik A. Richter2
1Institute of Inflammation Research, Department of Rheumatology, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark
2Department of Nutrition, Exercise and Sports, University of Copenhagen, Copenhagen, Denmark

Key points

• Exercise represents a challenge to diabetes regulation, particularly in type 1 diabetes.
• Based on experience and frequent blood glucose measurements, the insulin dosage must be adjusted to intensity and duration of exercise and food intake.
• The diabetes patient in good metabolic control should reduce the insulin dosage prior to exercise, more so the higher the energy expenditure during exercise.
• In youth prone to spontaneous exercise, carbohydrate administration may be a more feasible means of preventing hypoglycemia than adjustment of insulin dosage.
• Paying due respect to possible long-term complications, with proper education and medical control and guidance diabetes patients can participate in all kinds of sports, even at world-class levels.
• In type 2 diabetes, but not in type 1, a reduction in HbA1c can be expected with regular exercise.
• Regular exercise reduces the risk of cardiovascular morbidity and increases life expectancy in diabetes.
• Exercise reduces the risk of gestational diabetes and is also recommended for individuals with this condition.
• Physical activity reduces the risk of getting type 2 diabetes, especially in people at increased risk.

Introduction

A certain minimum of skeletal muscle activity in excess of that required for basal pulmonary ventilation is necessary for all people, including diabetic patients, for everyday activities such as dressing, cooking, and eating. House work and walking and bicycling for the purpose of transportation represent heavier exercise, and intense physical activity may be required for occupational activities in agriculture, forestry, fishing, industry and, obviously, professional sports and ballet. In addition to such obligatory exercise, many people also engage in sports to increase their physical, psychological, and social well-being.

However, the trend is that the advance of modern technology promotes physical inactivity. This is unfortunate, because physical activity has a pivotal role in health promotion and prevention of disease, including type 2 diabetes (T2DM) [1,2]. For that reason it has become necessary for national health authorities to recommend planned, regular exercise [3]. The recommendation also applies to diabetic patients. Among the weighty arguments are that evidence exists for both type 1 and type 2 diabetics that exercise reduces the risk of cardiovascular events [4,5] and increases life expectancy [6–8]. The recommendation for patients with diabetes to take physical activity is supported by both national academic, diabetes and sports associations, for example the American Diabetes Association and the American College of Sports Medicine [9].

Because exercise is an essential part of life and, furthermore, is associated with potential health benefits, it is important to be aware of the inherent risks and to know how it influences the body with regard to organ function, dietary needs, and interaction with medicine and the environment. This is of particular interest in people with diabetes since many of the characteristics and manifestations of this disease would be expected to interfere with work performance and vice versa.

Exercise physiology

In healthy humans who perform severe dynamic exercise with big muscle groups, oxygen uptake may increase 15–20-fold due to increased metabolic rate in the working muscles. In these, arterioles dilate, capillaries are recruited and fluid accumulates intra- and extracellularly. Pulmonary ventilation may exceed 100 L per min and cardiac output increases to 20–30 L per min to meet the needs of the working muscles, while blood flow to resting muscles, the splanchnic area, and the kidneys decreases.
Systolic and mean arterial blood pressure increase. Some of the produced heat is accumulated in the body, and the core temperature may exceed 40°C. Most of the heat is removed by evaporation, however, and the sweat rate during exercise in hot and humid air may be 2 liters per hour. The glomerular filtration rate decreases while potassium is released from contracting muscle and may attain a concentration in plasma of 8 mEq L$^{-1}$ at maximum exercise.

During exercise in the postabsorptive state the energy demand of the working muscles is met by breakdown of intramuscular glycogen and triglyceride stores as well as by an increased supply of glucose from the liver and free fatty acids (FFA) from extramuscular triglyceride stores [10]. Protein oxidation does not increase significantly unless muscle glycogen stores are exhausted. The relative contribution of carbohydrate to overall energy delivery increases with exercise intensity, while fat oxidation in absolute terms peaks at 70% of VO$_{2\text{max}}$ and becomes negligible at maximum exercise. Breakdown of intramuscular energy stores as well as overall carbohydrate oxidation wanes with time during mild to moderate exercise. The increase in glucose production in response to exercise primarily reflects hepatic glycogenolysis but contribution from gluconeogenesis increases with duration of mild exercise. At moderate workloads the plasma glucose concentration is usually well maintained (Figure 41.1). However, at high work intensities the glucose concentration increases [10,11],

![Figure 41.1](image-url)

*Figure 41.1* Simplified schemes of substrate fluxes and their hormonal regulation in healthy subjects in the postabsorptive state at rest (upper panel) and during moderate exercise (lower panel). FFA means free fatty acids; “+” and “−” indicate enhancing and inhibiting effects, respectively. During exercise a decrease in plasma insulin and an increase in concentrations of counterregulatory hormones are seen, while plasma glucose remains constant, because an increase in glucose uptake in contracting muscles is matched by an increase in hepatic glucose output.
while hypoglycemia may develop during prolonged exercise [12], these deviations being favored when prior to exercise carbohydrate stores are high and low, respectively [13,14]. Because selection of fuel for oxidation is affected by substrate availability [10,15], exogenous nutrients, particularly glucose, influence muscle metabolism when administered before or during exercise [16–18].

**Hormonal regulation**

The autonomic neuroendocrine system serves the coordination of the various functions of the body to allow coping with the physical challenges of life. Accordingly, it follows from the depicted multiple perturbations that depending on its intensity and duration, exercise may elicit extensive autonomic endocrine responses. These adaptations resemble those seen in hypoglycemia, early fasting, hyperthermia, surgery, and hemorrhage, that is, they include a decrease in plasma insulin as well as an increase in sympathetic nervous activity, insulin counterregulatory hormones, and hormones retaining sodium and water in the kidneys [15]. The endocrine adaptations to exercise are based on the same fundamental regulatory principles that are thought to control circulatory, respiratory, and thermoregulatory responses in exercise, a fact guaranteeing that the responses of the major physiologic systems are highly coordinated [15]. At the onset of exercise, nervous impulses from motor centers in the brain and from working muscles elicit an increase in sympatho-adrenal activity and in the release of some pituitary hormones (e.g. GH, ACTH, β-endorphin, prolactin, and ADH). This exercise-specific, “feedforward” stimulus is determined by workload relative to the individual's maximum capacity (V_{O2}/V_{O2max}) rather than by absolute workload. The primary neuroendocrine responses, in turn, change the secretion of subordinate endocrine cells, for example sympathetic nervous activity depresses insulin secretion by α-receptor-mediated mechanisms, while the renin-angiotensin-ADH system is stimulated by β-receptor mechanisms, and ACTH stimulates cortisol secretion.

The fast nervous mechanisms intimately related to contractile activity operate throughout exercise, but the hormonal responses may be gradually intensified according to feedback from metabolic error signals (among which a decrease in glucose availability is the most important) as well as from nonmetabolic error signals sensed by blood pressure and volume, osmolality, electrolyte and temperature receptors. At high work rates, feedforward control dominates. At low exercise intensities, just as at rest, feedback regulation dominates the hormonal changes, and at least in humans, an increase in plasma glucagon concentration is always predominantly determined by a decrease in plasma glucose availability. The primary setting of the hormonal response to exercise depends on the state of the organism with respect to training, nutrition, hydration, phase of menstrual cycle, and state of health [15].

**Hormonal influence on exercise metabolism**

The overall hormonal response to exercise appears to be geared to stimulation of catabolic processes resulting in lipolysis and glycogenolysis in adipose tissue, muscle, and liver. With regard to hepatic glucose production during exercise, the regulation particularly in humans is not clear [15]. A decrease in plasma insulin concentration is always essential, because glucose production during exercise varies inversely with modest artificial changes in insulin levels [15]. In rats and dogs the exercise-induced increases in glucagon and epinephrine concentrations are the major determinants of the increase in glucose production [10,15]. Furthermore, the potency of the rise in glucagon has been shown to be enhanced by the accompanying fall in insulin [19]. On the other hand, both in these species, and in guinea pigs [20] and humans [21,22] sympathetic liver nerves do not influence glucose production in exercise. Strangely, neither epinephrine [22,23] nor glucagon [10,22] or angiotensin [24] seem to be major stimuli in humans. In healthy man, arterial glucagon does not increase above basal levels until after more than 1 hour of submaximal exercise, and it even decreases during maximal exercise, when glucose production peaks [10].

It may seem surprising that during exercise the glucose uptake in muscle varies inversely with the rate of secretion and the plasma concentration of insulin [10]. However, muscle contractions per se enhance muscle glucose uptake by mechanisms that are independent of insulin [25,26]. These mechanisms include feedforward components, which are probably Ca^{++} mediated [27,28], as well as feedback components, which may be mediated by AMPK [29–31]. Recently the involvement of the small cytoskeleton-regulating GTPase Rac1 has also been demonstrated [32]. In isolated skeletal muscle, contractions are a more powerful stimulus to glucose transport than insulin, and the two stimuli are fully additive [33]. They translocate glucose transporting protein (GLUT4) to the sarcolemma by at least partly differing signaling mechanisms and have different preferences for recruitment from the various intracellular GLUT4 stores [34]. During exercise the decrease in plasma insulin curbs the increase in glucose uptake in muscle, and the same is true for the increase in plasma epinephrine [23]. The effect is due to the direct actions of hormonal changes on glucose transport as well as to the inhibition of glucose uptake secondary to an enhancement of lipolysis and muscle glycogen breakdown [10,35,36]. In addition to the decrease in plasma insulin, both sympathetic nerve activity and epinephrine are major determinants of adipose tissue lipolysis in exercise [37].

The primary setting of substrate mobilization in exercise is simple. The same enzymes, that is, glycogen phosphorylase and hormone-sensitive lipase (HSL) and adipose tissue triglyceride lipase (ATGL), are flux generating for glycogenolysis and lipolysis, respectively, in both muscle and extramuscular stores of glycogen (i.e., liver) and triglyceride (i.e., adipose
tissue). Furthermore, the enzymes are stimulated by the same neuroendocrine changes and in muscle they are also stimulated in parallel by mechanisms accompanying contractions [36,38,39]. The autonomic neuroendocrine response to exercise includes increases in several components potentially enhancing the activity of glycogen phosphorylase, HSL and ATGL, and, in addition, a decrease in insulin, a hormone inhibiting these enzymes. Metaphorically speaking, substrate mobilization in exercise is increased by stepping on the accelerator while releasing the brake (Figure 41.1).

**Metabolism during exercise in diabetics**

**Type 1**

While the insulin concentration in the plasma of healthy people is automatically adjusted to any work condition, this adjustment does not take place in type 1 diabetics. Instead they have to carry out the work during the influence of the given amount of insulin they have administered subcutaneously. Therefore, the changes in metabolism that take place in type 1 diabetics in response to exercise usually reflect either too much or too little insulin in plasma.

While in healthy people plasma insulin is low during exercise, that of diabetics is not low if they take the same dose of insulin as they use on non-exercise days. In diabetics, in contrast to healthy subjects, plasma insulin may even increase during exercise because absorption of insulin may be favored by exercise-induced movements of the site of administration in the skin and increases in skin blood flow [40,41]. A high plasma insulin level implies risk of hypoglycemia because the normal increase in hepatic glucose production is inhibited, and insulin stimulates muscular glucose uptake both directly and by an inhibition of FFA release from triglyceride stores and, in turn, of fat combustion [40] (Figure 41.2).

If, then, type 1 diabetics reduce their dose of insulin, they risk getting too little insulin in plasma [42]. In this event, hepatic glucose production and the release of FFA from adipose tissue increase more than normal, while muscular glucose uptake is inhibited (Figure 41.2). The high plasma FFA level and an increased hepatic fractional extraction and conversion to ketone bodies of FFA result in an increased production of ketone acids and increased muscular fat oxidation [43]. In patients who are insulin-deficient prior to exercise, increased fat oxidation during work reflects increased stores of intramuscular triglycerides and these stores will be depleted during exercise [44–46]. In such patients the muscular glycogen concentration and exercise-induced glycogen depletion are reduced in comparison to those of healthy subjects [45]. Nevertheless, lactate accumulation is enhanced as a consequence of incomplete carbohydrate oxidation due to increased fat oxidation and diminished pyruvate dehydrogenase activity [44,47,48]. Furthermore, an increased production of lactic acid, glycerol, and glycogenic amino acids results in higher hepatic gluconeogenesis during exercise than in control subjects [44,47]. Taken together, in insulin-deficient exercising individuals, the plasma glucose concentration increases and acid accumulates—the diabetic state deteriorates [42,47].

It is evident that the physically active diabetic, like the sedentary one, balances between hyper- and hypoglycemia. However, exercise accelerates the rate at which both of these unpleasant alternatives develop (Figure 41.3). Exercise increases sympathetic nervous activity and the plasma concentrations of counterregulatory hormones, a phenomenon which per se causes an accumulation of glucose and lactic acid and ketone bodies in plasma by enhancing glucose and FFA production and inhibiting glucose uptake in muscle [49,50]. Moreover, in diabetics who have taken too little insulin the autonomic neuroendocrine response is exaggerated, and so are its effects. Therefore, these patients are especially prone to hyperglycemia and ketoacidosis [43,48,51–54]. The increase in plasma glucose during exercise may even in itself stimulate the counterregulatory hormonal response [54]. In other words, the type 1 diabetic who exercises with a plasma insulin concentration that is too low, in terms of fuel mobilization steps harder on the accelerator and less hard on the brake than healthy subjects do.

On the other hand, the plasma glucose concentration decreases more rapidly in diabetics, who take too much insulin, if they also exercise [55–57] (Figures 41.2 and 41.4). This phenomenon is due to the fact that glucose uptake in muscle becomes extra high because contractions per se increase glucose uptake in muscle cells and also increase the sensitivity to insulin of glucose uptake in these cells [58] and because delivery of insulin and glucose to the sarcolemma of contracting muscle cells is increased by increases in blood flow and capillary permeability surface area product in exercising muscle [59]. Furthermore, the indirect enhancing effect of insulin on glucose uptake in muscle, which is exerted via inhibition of fat cell lipolysis, may be augmented in exercise compared with rest, because insulin prevents the marked increase in lipolysis which normally occurs during exercise. In addition, insulin can prevent the rise in hepatic glucose production during exercise. Because glucose turnover is accelerated during exercise, inhibition of hepatic glucose production leads to more marked hypoglycemia than at rest when glucose turnover is less. Of note are recent studies indicating that in type 1 diabetics as well as in healthy subjects antecedent hypoglycemia or exercise blunts neuroendocrine and metabolic counterregulatory responses to subsequent hypoglycemia or exercise, and, accordingly, increase the risk of severe hypoglycemia at either stress [60].

**Type 2**

Type 2 diabetics, in contrast to type 1 diabetics, have some insulin secretory capacity, and their insulin secretion changes with needs in principle as it does in normal individuals. Accordingly, if they are not treated with exogenous insulin, they cannot respond as dramatically to exercise as type 1 diabetics. In other words, they do not run the same metabolic
Figure 41.2. Simplified schemes of substrate fluxes and their hormonal regulation during moderate exercise in type 1 diabetes patients, who have taken a normal dose of insulin (upper panel) or are insulin deficient (lower panel). After insulin administration the uptake of glucose in contracting muscles increases more than the hepatic glucose output, and the plasma glucose concentration falls. In contrast, without insulin the plasma glucose concentration rises, because hepatic glucose output exceeds muscle glucose uptake. The increases in counterregulatory hormones and FFA are exaggerated and the production of ketones is increased.

risks with exercise. Still, although insulin resistance is a major hallmark of T2DM contraction-mediated glucose transport in type 2 diabetic muscle is not reduced [59,61,62]. It has been claimed to rely more on NO than seen in normal muscle [63], but exercise-induced increases in atypical PKC and AMPK alpha2 activities in muscle are normal [64,65].

So, in diet-treated type 2 diabetics, who perform prolonged mild to moderate exercise in the postabsorptive state, a marked reduction in hyperglycemia can be seen [66,67]. This reflects that hyperglycemia enhances the exercise-induced increase in muscle glucose uptake and inhibits the increase in hepatic glucose production, effects which are reinforced by the fact that the secretion and, in turn, plasma concentration of insulin, at least as long as hyperglycemia prevails, are not depressed as readily as in healthy subjects [66,68–71].

The rate of decrease in plasma glucose concentration is augmented, if sulfonylurea drugs are taken shortly before exercise (Figure 41.5) [72]. Sulfonylureas stimulate insulin secretion, and although the secretion is depressed by exercise, the plasma insulin concentration remains higher than during exercise without prior drug intake (Figure 41.5). This relative hyperinsulinemia explains that the exercise-induced rise in hepatic glucose production is lower and, accordingly, the hypoglycemic effect of exercise is larger than when exercise
Figure 41.3 An illustration of the delicate balance between too little and too much insulin available during exercise. After an overnight fast 24-year-old type 1 diabetes patients injected 4 IU of regular insulin either 3 h (white dots, \( n = 11 \)) or 1 h (black dots, \( n = 15 \)) before intermittent, graded, exhausting ergometer cycle exercise (5 × 10 min at 50–150 W with 5-min rest intervals). The plasma glucose concentration increased 5.1 mM and decreased 4.7 mM, respectively. The broken line represents data from six healthy subjects. Source: Adapted from Zander 1983 [42]. Reproduced with permission of Thieme Publishing Group.

is performed without acute sulfonylurea intake. Interestingly, the hyperinsulinemia apparently overrides the effect of the augmented counterregulatory hormone response observed when sulfonylureas are taken (Figure 41.5) [72]. Abstaining from sulfonylurea treatment for 5 days does not result in an exaggerated counterregulatory hormone response and hyperglycemic reaction to exercise as seen in type 1 diabetics after insulin withdrawal [73]. In contrast, during 60 min of moderate exercise the plasma glucose concentration decreases more after prolonged sulfonylurea abstention than after abstention only from the morning dose prior to exercise [73]. This is because the basal plasma glucose level increases during the prolonged drug abstention, and, in turn, maintains the basal insulin secretion and impairs the exercise-induced depression of insulin secretion. The augmented hyperglycemia and relatively high insulin levels then blunt the exercise-induced increase in hepatic glucose output, thereby promoting a decrease in plasma glucose [73]. Still, an exaggerated counterregulatory hormone response and glucose production and an ensuing increase in plasma glucose is seen in type 2 diabetics performing brief intense exercise in the postabsorptive state [74].

The plasma glucose concentration also decreases markedly during prolonged moderate exercise in type 2 diabetics in the postprandial state (Figure 41.6) [75]. This is accompanied by
Figure 41.4 The interaction between breakfast (B), ergometer cycle exercise (45 min at 55% V_{02max}) and different doses of insulin (I) in 15–18-year-old type 1 diabetes patients treated with multiple subcutaneous insulin injections (MSI, upper panel, n = 6) or continuous subcutaneous insulin infusion (CSII (pump), lower panel, n = 7). If premeal insulin dose is reduced plasma glucose concentration increases before exercise but during exercise hypoglycemia is avoided. The two regimens allow similar coping with exercise. *p < 0.05 and **p < 0.01 vs. rest. Source: Adapted from Schiffrin [57]. Reproduced with permission of American Diabetes Association.

Reduced meal-induced increases in insulin and C-peptide concentrations and reflects an exercise-induced increase in glucose clearance, while glucose appearance in plasma is unchanged [75]. The plasma glucose lowering effect of exercise is higher after a meal than in the fasted state, even at a time at which only plasma insulin and not plasma glucose is increased in the former compared to the latter condition [76]. In contrast to findings in the postabsorptive state [74], in the postprandial state hyperglycemia is not aggravated during severe exercise [77]. This probably reflects that the counterregulatory hormonal response to exercise is lower in the latter compared with the former state, and that, furthermore, the high glucose and insulin concentrations accompanying food intake inhibit an exercise-induced increase in hepatic glucose production and enhance an increase in glucose uptake in muscle [77]. Repeated, high intensity exercise in fact reduces both glucose concentrations and insulin secretion after a meal [77]. Interestingly, exercise-induced reductions in glucose, insulin, and C-peptide responses are similar during repeated, high intensity exercise compared with isocaloric but prolonged moderate postprandial exercise [77]. This indicates that in the postprandial state overall energy expenditure is the major determinant of exercise-induced changes in overall glucose homeostasis and insulin secretion in type 2 diabetic patients. The peak exercise intensity is not of major importance in this respect. Apparently, in terms of reducing glucose levels and insulin
secretion following a meal, acute exercise has little effect beyond that of increasing energy expenditure. In line with this view, it has been shown that a reduction of energy content of a meal has the same lowering effect on postprandial glycemia and insulin secretion as an equivalent exercise-induced increase in energy expenditure (Figure 41.6) [75]. Neither prolonged moderate nor intermittent severe exercise performed in the postprandial state influence glucose and insulin responses to a subsequent meal [75,77]. Besides postprandial hyperglycemia, the exaggerated meal-induced increases in chylomicron and very low-density lipoprotein triglyceride concentrations are also reduced by postprandial prolonged exercise in type 2 diabetics [78].

**Special patient groups**

**Children**

$V_{02\text{max}}$ and physical work capacity are impaired in young people with T1DM [79]. Inactivity likely contributes to lower fitness and performance because when matched for age, body size, and physical activity, adolescents with T1DM have similar cardiac function and $V_{02\text{max}}$ to healthy adolescents [79]. Children often engage in spontaneous exercise the extent of which is generally unpredictable, a fact which makes appropriate adjustments in insulin and diet difficult. Such adjustments are also rendered difficult by the fact that during similar conditions the variation in glycemic response to exercise is high in children both on the inter- and intra-individual level [80]. Furthermore, glycemic awareness seems to be poor in adolescents with T1DM [81], and their perceived exertion is high compared to findings in control subjects [82], while carbohydrate utilization may be impaired [83]. Hypoglycemia during prolonged moderate postprandial exercise is a particular threat due to prevailing hyperinsulinemia, but intermittent ingestion of carbohydrate beverages matched with total carbohydrate utilization has been shown to markedly attenuate the drop in plasma glucose in exercising adolescents with T1DM [84]. The fact that severe postexercise late-onset hypoglycemia may be particularly prevalent in active diabetic children probably reflects that proper strategies are not adopted to replace muscle and liver glycogen stores [85]. If, on the other hand, brief high intensity exercise is performed, a dramatic increase in plasma glucose along with marked increases in catecholamine concentrations may be seen even in well-insulinized diabetic children [86], just as is the case in diabetic adults [87]. Because requirements for insulin and food change with growth and sexual maturation, physical
activity in youth with diabetes has to be particularly carefully managed. In the case of an amenorrheic athletic girl with T1DM, dietary instruction is paramount as part of the exercise program, not only as treatment of the exercise-induced menstrual disorder, but also to maintain as near-normal plasma glucose levels as is safe and feasible, because chronic hyperglycemia delays puberty and is associated with short stature [88]. Today almost 50% of children and adolescents under the age of 18 with diabetes have T2DM, and 80% of these are overweight [89]. These facts call for attention with regard to physical activity in the prevention and treatment of T2DM in children [89].

Pregnant women

Oligo- or amenorrhea may in both healthy and diabetic women result from excessive physical training [15]. If these women want to become pregnant they must reduce or abstain from exercise. Even in normally menstruating insulin-treated women this step may be necessary to achieve a glucoregulation compatible with fertilization, because exercise adds an additional variable that may make glucose control difficult [88]. For the same reason it has been concluded that the utility of an exercise program as part of the glucose control program for pregnant women with pregestational diabetes, be it type 1 or type 2, is questionable [88]. The paramount management goal for the best outcome of pregnancy is to achieve and maintain normoglycemia. However, exercise-induced deterioration may only be associated with a risk of spontaneous abortion if plasma glucose levels exceed 4 standard deviations of the mean level of a healthy pregnant woman [88].

Insulin sensitivity declines progressively during the second half of gestation, and if insulin secretion is not increased appropriately, the previously nondiabetic pregnant woman develops gestational diabetes. However, physical activity before and during pregnancy reduces the risk of both gestational diabetes and pre-eclampsia [90]. Exercise of moderate intensity is recommended along with diet in the treatment of patients with gestational diabetes, because exercise may lower plasma glucose levels and insulin needs [91–93] and reduces weight gain [94] in the patients.

After menopause basal metabolic rate decreases. Exercise may help postmenopausal women with T2DM to avoid weight gain [88]. If postmenopausal women with diabetes undertake hormonal replacement therapy with estrogen and progesterone their insulin requirement increases. When an exercise program is combined with hormonal replacement therapy, the insulin regimen becomes especially challenging [88].

Autonomic neuropathy

Diabetic autonomic neuropathy influences cardiovascular and hormonal responses to exercise. Exercise endurance as well as VO2max are diminished by autonomic neuropathy [95,96]. Vagal insufficiency results in an increased heart rate at rest, and impairment of both parasympathetic and sympathetic nervous function results in diminished increases in heart rate in response to exercise [95–98]. In patients with diabetic autonomic neuropathy stroke volume is diminished both at rest and during exercise [95]. This seems to reflect impaired cardiac contractility, which may be due to diminished cardiac adrenergic stimulation, impaired sympathetically mediated dilatation of coronary resistance vessels, and diabetic cardiomyopathy [95,99]. The mentioned changes imply that in diabetics with autonomic neuropathy cardiac output increases less with exercise than in diabetics without neuropathy [95]. Further, probably reflecting splanchnic sympathetic neuropathy and, in turn, reduced splanchnic vasoconstriction, the exercise-induced increase in hepatosplanchnic vascular resistance is also diminished in patients with autonomic neuropathy [95]. Defective increases in cardiac output and hepatosplanchnic resistance explain that the arterial blood pressure response to exercise
may also be reduced [95,97]. The mentioned impairments in cardiovascular function are less pronounced in patients with diminished beat to beat variation in heart rate during deep breathing than in patients, who also have orthostatic hypotension, showing that the impairments are directly related to the degree of autonomic neuropathy [95,96]. In patients with autonomic neuropathy the hypoxic ventilatory drive may be impaired [100], but nevertheless these patients may show an exaggerated pulmonary ventilation during heavy exercise [101]. Impaired sympathetic function is accompanied by diminished catecholamine responses to exercise, and responses of GH, cortisol, renin, and pancreatic polypeptide may also be diminished in diabetics with autonomic neuropathy [102–104]. However, metabolite responses may be only slightly influenced, a fact which probably reflects augmented sensitivity of target tissues to hormonal stimulation in these patients [105].

**Adaptations to training**

**Insulin action and secretion in healthy subjects**

Insulin action on whole body glucose disposal is reduced in the early recovery period after short-term maximal dynamic exercise in both trained and untrained subjects [11], and insulin action is also reduced after marathon running [106]. These findings have been ascribed to elevated levels of counterregulatory hormones as well as metabolic inhibition of glucose uptake in muscle by accumulated glycolytic intermediates and increased fat metabolism, respectively. After a marathon the reduced insulin action is accompanied by impaired muscle glycogen resynthesis in the face of unaltered GLUT4 content and increased glycogen synthase activity [106,107]. Unaccustomed, severe exercise, for example eccentric exercise, which causes muscle damage also impairs insulin action [108,109]. The reduced insulin action after eccentric exercise includes diminished insulin action on glucose uptake and glycogen synthesis in the recruited muscles and is seen in the face of diminished total GLUT4 content [110], decreased insulin signaling [111] but unaltered glycogen synthase activity [108,109]. In accordance with these findings insulin-mediated glucose transport has been shown to be decreased in rat muscle after eccentric contractions, while glycogen synthase responded normally [112].

However, nonexhausting endurance-type exercise that consists predominantly of concentric contractions enhances the effect of insulin on glucose disposal [113–115], and insulin action increases over the full range of activity levels from inactivity during bedrest to daily endurance training [109,116]. Also regular strength training enhances insulin action [109]. The exercise-induced increase in insulin effect is predominantly located in the recruited muscles and reflects local contraction-dependent mechanisms [109,115]. The enhancing effect of acute exercise to a large extent reflects muscle glycogen depletion and accompanying glycogen synthase activation and includes an increase in insulin sensitivity of muscle glucose transport [109,113,117]. With increasing habitual physical activity level the muscle GLUT4 content increases and so does insulin responsiveness of glucose transport [109]. Also muscle capillarization increases, and these changes may to some extent reflect conversion of muscle fibers from type II B/X to type II A and perhaps further on to type I [109]. The muscle content of key enzymes in glucose storage and metabolism also increases with level of physical activity [109,118]. The mentioned changes explain that muscle glucose extraction during insulin stimulation varies directly with activity level [109]. Because the ability of insulin to stimulate muscle blood flow also increases with activity level, insulin-mediated glucose uptake in muscle must vary directly with level of habitual activity [109]. After acute exercise glucose taken up in muscle is predominantly subjected to storage as glycogen while glycolysis is diminished [109]. In contrast, in the trained as well as in the inactive state insulin-mediated glycogen storage and glycolysis essentially vary in parallel [59,109].

Exercise may also enhance insulin action in extramuscular tissues. The hepatic glucose storage capacity is increased by training [109]. Furthermore, in trained subjects the suppressive effect of insulin on hepatic glucose production is increased probably reflecting an effect of the last exercise bout [109,118]. The suppressive action of insulin on pancreatic secretion of insulin and glucagon may also be enhanced by exercise making it difficult to tell whether exercise-induced changes in hepatic sensitivity to insulin are direct or secondary to changes in glucagon secretion [109]. In line with the view that the insulin sparing effect of training is more pronounced in peripheral tissues than in liver, studies carried out during ordinary living conditions have shown that a relatively smaller fraction of secreted insulin is delivered to "posthepatic" tissues in athletes compared to healthy subjects [109]. This finding also points at a higher hepatic insulin extraction after training [109]. Microdialysis studies have shown that training enhances insulin-mediated glucose uptake in human subcutaneous adipose tissue [119], and studies in rats have shown a similar effect in several intra-abdominal adipose tissues [120]. In line with these findings, in adipocytes, insulin-mediated glucose transport as well as GLUT4 mRNA and protein are increased by training [109]. While such effects are not elicited by a single bout of exercise, insulin-mediated suppression of lipolysis is enhanced by both acute exercise and training [109,119]. Training decreases basal concentrations in plasma of triglyceride, total cholesterol and LDL cholesterol, and increases the HDL fraction of total cholesterol [121]. Postprandial lipidemia is also reduced by both acute exercise and training. The effect of training probably reflects increased lipoprotein lipase activity in muscle [109,122]. In addition to enhancing the effect of insulin on carbohydrate and fat metabolism, physical activity also enhances the protein anabolic effect of insulin in human muscle [109,123,124].
Adaptations to training in diabetics

Type 1

If type 1 diabetics have a normal activity level and do not have long-term complications they also have a normal oxidative capacity of skeletal muscle and normal physical work capacity [121]. However, their skeletal muscles are insulin resistant [121]. These patients generally respond to both endurance and strength training just as healthy subjects, for example by diminished heart rate and blood pressure during rest and exercise, and increased muscle force, respectively [40,121,127]. Insulin sensitivity increases and the activity of oxidative enzymes, glycogen synthase, and hexokinase increases in muscle [40,79,121,128]. Also muscle capillarization increases but the increase may be less pronounced than in healthy subjects [40,121]. A case report indicates that improvement in insulin sensitivity through physical training may help to induce and prolong remission at onset of T1DM [129]. However, after the honeymoon period a decrease in insulin requirements with training is not a universal finding, possibly because carbohydrate intake may increase [40,121]. Since exercise makes diabetes more difficult to control, the daily fluctuations in blood glucose may increase, and the average plasma glucose concentration, which is reflected in the HbA1c level, does not generally decline during training [40,79,121]. An improved glucose regulation is more likely to occur if patients have been poorly controlled prior to training [121], probably because training is then accompanied by more attention to disease control [79]. However, type 1 diabetic patients who participate in competitive sports may even experience deterioration of glycemic control in spite of improvements in fitness [130]. This probably reflects an irregular schedule, high exercise intensities, and measures to avoid hypoglycemia in response to exercise, that is, excessive reductions in insulin dose and excessive intake of carbohydrate. Additionally, these patients may not show increased insulin sensitivity [130]. Probably reflecting the variation seen in effect of training on metabolic homeostasis some studies have found changes in plasma triglyceride and cholesterol levels upon training similar to those seen in healthy subjects, whereas other studies have not [40,121].

Type 2

Patients with T2DM have a low VO2max, a low maximal peripheral O2 extraction, and a reduced fraction of oxidative, highly insulin-responsive muscle fibers compared with healthy subjects even when matched for levels of ambient activity [118,131–133]. In agreement with this, their nonobese, nondiabetic first-degree relatives also have a low level of fitness and a high content of nonoxidative fibers in skeletal muscle [109,134]. The diminished VO2max in type 2 diabetics may be explained by myocardial dysfunction and by impaired oxygen extraction due to mitochondrial dysfunction and impaired increase in capillary surface area in working muscle [73]. The myocardial phosphocreatine (PCr)/ATP ratio may be diminished in resting
diabetics, while during exercise an attenuated increase in blood flow \[73\] is accompanied by a faster decrease in PCr and pH in skeletal muscle \[135\].

In type 2 diabetics \(V_{\text{O2max}}\), as well as the enzymatic potential for oxidation and fat combustion in muscle increase normally during endurance training \[40,121,136,137\], whereas the increase in capillarization in muscle is impeded \[40,121\]. Bedrest exacerbates the insulin resistance of obese individuals with pre-existing impaired glucose tolerance \[118\]. Conversely, in the postabsorptive state a single bout of prolonged sub-maximal dynamic \[138\] or resistance \[139\] exercise improves whole-body insulin sensitivity in type 2 diabetics. The effect of dynamic exercise has been shown to be located both in liver and peripheral tissues \[138,140\]. Impairment of the activity of atypical protein kinase C in muscle is probably important for the reduced insulin-mediated glucose transport in T2DM, and this activity is increased by exercise \[64\]. In contrast to the reaction to submaximal exercise, in response to brief intense dynamic exercise insulin sensitivity is reduced for the first 1–2 hours postexercise but increased after 24 hours \[74\]. Also in the postprandial state moderate exercise seems to increase insulin-mediated glucose disposal in the immediate postexercise period, but the effect does not persist during or after the next meal \[75\].

Regularly repeated exercise may increase insulin sensitivity in both elderly glucose intolerant subjects \[126\], women with gestational diabetes \[92\], and type 2 diabetic patients \[136,141\] and their first-degree relatives \[109,134\] more than can be inferred from a single bout of exercise. However, the improvement may in type 2 diabetics be difficult to detect on the whole-body level \[118,142\]. Studies in which type 2 diabetic patients had catheterization of leg vessels before and after bicycle training have shown that the effect of training is predominantly, but not entirely, due to an enhanced insulin-mediated glucose clearance in skeletal muscle (Figure 41.7) \[136\]. Possibly, an increase in the activity of the transcriptional co-activator PGC-1α and accompanying amelioration of mitochondrial dysfunction and, in turn improved oxidative metabolism, at least in part account for enhanced insulin action in muscle \[143,144\]. In terms of insulin signaling, it has recently been shown that the TBC1 domain family of GTPases are important regulators of GLUT4 translocation \[145\] and endurance training has been shown to be able to restore impaired site-specific phosphorylation of TBC1 domain family, member 4 (TBC1D4) in skeletal muscle of T2DM patients along with improvement in insulin sensitivity \[146\]. Insulin-mediated increase in blood flow and glucose extraction in muscle are augmented following training \[134,136\]. The increase in extraction is in line with a training-induced increase in muscle GLUT4 protein and mRNA \[147\]. The increase in glucose clearance reflects increases in both nonoxidative glycolysis and glucose storage and parts of the insulin signaling pathway \[136,146\]. Increases in fractional velocity of glycogen synthase and mRNA for this enzyme have been found after endurance training in diabetic muscle but total glycogen synthase, hexokinase, and lactate dehydrogenase activities were unaltered \[136\]. Interestingly, in T2DM patients this type of training may increase muscle glucose clearance to values seen in healthy untrained control subjects.
Strength training has also been shown to increase insulin action on glucose clearance in skeletal muscle of T2DM patients [148,149]. The effect was independent of changes in muscle mass and reflected an enhanced insulin-mediated increase in muscle blood flow as well as increases in protein content of the insulin receptor, GLUT4, protein kinase B α/β (Akt 1/2) and glycogen synthase.

In addition to the described effects that may be seen in weight-stable patients [136], training may increase diet-induced weight loss, and this occurrence may, in itself, increase insulin sensitivity [109,150]. Regular exercise causes a loss of intra-abdominal adipose tissue, and this correlates with improvement in insulin action [141]. However, training per se may not generally result in lower body mass, a fact which may be explained by either changes in opposite directions of lean and fat body masses, respectively, or by reduced physical activity between training periods, or by increased food intake [151].

A training-induced reduction in glucose-stimulated β-cell secretion as seen in healthy subjects would appear less expedient in T2DM patients. However, some studies have, in fact, indicated improvement in β-cell function by training in these patients [40,109]. In a study using stimulation with hyperglycemic clamping and arginine bolus, increases in β-cell responses were seen in patients, who prior to training had a moderate residual β-cell function, whereas no change was seen in patients with a low remaining secretory capacity [109]. One possible explanation could be that a diminished glucose stress resulting from an increase in target tissue insulin sensitivity may improve the secretory capacity of overloaded, yet recoverable, β cells. However, in the mentioned hyperglycemic clamp study no change in insulin sensitivity and HbA1c was seen. Interestingly, in agreement with the findings in humans, training has been shown to improve glucose-stimulated insulin secretion in rats made mildly or moderately diabetic by partial pancreatectomy, but training did not influence secretion in more severely diabetic rats [152]. Also in that study other mechanisms than reduced glycemia apparently accounted for the positive β-cell adaptation to training [152].

The impact of the training-induced adaptations in both secretion and effect of insulin on glucose homeostasis during ordinary living conditions has been evaluated in a study in which type 2 diabetic patients had blood samples drawn around the clock. Areas under plasma concentration versus time curves were identical for glucose while energy intake was at least as high when compared to before training [109]. Integrated insulin and C-peptide levels were reduced after training indicating that during physiologic conditions insulin action is enhanced and insulin secretion spared by training [109]. The lack of change in integrated glucose levels is in accordance with the fact that a significant change in HbA1c concentration is often not seen in training studies. However, meta-analyses have revealed that training may reduce HbA1c by 0.66% [151], and that exercise intensity is a better predictor of improvement in HbA1c than exercise volume is [137]. The improvement of glucose homeostasis by regular exercise may be more marked in younger patients with T2DM and in patients with short duration of disease [109].

A single bout of prolonged dynamic exercise reduces postprandial chylomicron triglyceridemia in type 2 diabetics. Furthermore, regular exercise has been shown to reduce fasting triglyceride concentrations in these patients, whereas total cholesterol and its subfractions are not consistently changed [40,131,142,153,154].

Benefits, risks, and precautions associated with exercise in diabetes

Benefits

In T2DM training may improve overall glycemic control. In a meta-analysis of training effects a mean fall in HbA1c concentration of 0.66% was found and such an effect should be sufficient to markedly reduce the risk of diabetic complications and death [151,155,156]. Traditionally, endurance exercise has been recommended for patients with T2DM but recently it has been shown that a combination of endurance and strength training improves the metabolic outcome as evaluated by HbA1c levels [157]. The underlying mechanism is a training-induced increase in insulin sensitivity which is accompanied by beneficial relief of overloaded β cells and diminished plasma insulin levels. Adding to the risk reduction, training also tends to favorably influence body fat [158], arterial blood pressure [159,160], baroreflex sensitivity [161], endothelium-dependent vasodilation in both conduit and resistance vessels [134,162], skin blood flow [163], the heart rate × blood pressure product at rest and during exercise [158], hypertriglyceridemia and fibrinolysis [159,160,164], and has antioxidant and anti-inflammatory effects [165]. In line with these findings, cohort studies have in fact shown that in type 2 diabetic patients cardiovascular morbidity and mortality vary inversely with both physical fitness and level of physical activity and may be 50% lower in physically active compared with sedentary individuals [4,6,166].

In type 1 diabetics, exercise renders metabolic regulation difficult, and improvement of glycemic control should not be the aim of training [167]. However, endurance training in patients with T1DM has been shown to improve vascular endothelial function [168]. Furthermore, both epidemiologic prospective and retrospective studies indicate that regular physical activity in type 1 diabetics does protect against macrovascular as well as microvascular complications and increases life expectancy [5,7,168]. In both types of diabetes, additional proven benefits of regular physical activity are maintenance or improvement of
aerobic exercise capacity and muscle strength and, accordingly, a better tolerance towards physical demands. Another benefit that has been shown in healthy subjects and which in all probability also applies to diabetics is protection against osteoporosis. Furthermore, it is a general experience that exercise and sports in diabetics like in healthy subjects may improve mental health (e.g. reducing depression and anxiety, and enhancing well-being and self-esteem) and may promote social interaction and joy in a recreational environment [79,169].

**Metabolic risks and precautions**

In T1DM patients the immediate risk inherent in exercise is development of hyperglycemia and ketosis or hypoglycemia, respectively, during exercise. Both conditions will extend into the post-exercise period. Hypoglycemia may even develop many hours after exercise, because enhancement of insulin sensitivity and replenishment of glycogen stores from plasma glucose may last more than 12 hours [170,171]. This means that particular caution has to be exerted when exercise is carried out in the evening. To meet these challenges the diabetic has to be in good metabolic control prior to exercise [121,131]. Furthermore, the patient must imitate the normal response to exercise by reducing the insulin dose and/or increasing carbohydrate intake (Tables 41.1 and 41.2). The necessary adjustments depend on the exercise taken, the environment, and the individual, and, accordingly, have to be individually tailored based on frequent blood glucose measurements before, during, and after exercise. Extensive exercise should not be carried out earlier than 1–2 hours after a main meal and accompanying insulin administration. The short-acting monomeric insulin analogues are absorbed faster than regular insulin and, accordingly, the risk of undesirable interaction between peak insulin concentrations and exercise is over sooner after a meal if analogues are used compared with regular insulin [172,173]. The optimum insulin dosage depends on the intensity and duration of exercise and the more energy is spent the less insulin is needed [149]. In response to short-term strenuous exercise, transient hyperglycemia may develop in well controlled type 1 diabetic patients [87]. If the diabetic is not aware of the passing nature of the hyperglycemia, the patient may be tempted to inject extra insulin and may, in turn, provoke hypoglycemia [174]. For a given amount of work more insulin may be needed in the morning than in the evening, perhaps due to the higher plasma cortisol concentration in the morning. If the ambient temperature is high less insulin should be injected prior to exercise because absorption from the skin may be facilitated [174]. Neither in this nor in other situations can exercise-induced hypoglycemia be avoided by changing the insulin injection site [174].

Reductions in short-acting insulin dosage taken before exercise may amount to 30–80%, and postexercise reductions may also be necessary, for example to avoid nocturnal hypoglycemia [174,175]. Similarly, in order to perform long-term physical activity, patients treated with pump therapy may have to reduce the premeal insulin bolus by more than 50% and even to stop the basal infusion rate while exercising and to reduce it 25% for several hours afterwards [176,177]. On the other hand, a necessary reduction in insulin dosage prior to less extensive exercise may result in a greater increase in plasma glucose concentrations when a meal is taken after exercise than on non-exercise days when the full insulin dose is administered [40]. Accordingly, supplemental short-acting insulin may be required, particularly if food intake is increased after exercise. If plasma glucose is below 5.5 mM before physical activity, extra carbohydrate, for example a snack, should be eaten [121,131]. It is also recommended that during exercise sessions of moderate intensity lasting more than 30 min, 15–25 g extra readily absorbable carbohydrate should be taken by adult diabetics every 30 min [18]. Carbohydrate may be supplemented by 100 mL 5–8% sucrose solution every 10 min, whereby exercise-induced fluid losses are also substituted. After prolonged exhaustive exercise 100–200 g additional carbohydrate may be needed to optimize glycogen repletion and prevent late hypoglycemia [121,178]. Prevention of hypoglycemia by carbohydrate administration before, during, and after exercise may be more realistic, particularly in youth, than adjustment of insulin dose, because activities may be spontaneous and of unforeseeable intensity and duration [79]. The total amount of carbohydrate ingested should match the carbohydrate expenditure [84], and tables exist with estimates of carbohydrate utilization during various activities in children of various body masses [79]. However, it is of note that people with T1DM tend to overeat with exercise [178].

As in T1DM patients, exercise should also be avoided in T2DM patients if metabolic control is poor [121]. However, alarming exercise-induced hyperglycemia and ketosis has not been observed in T2DM patients, and neither has exercise-induced hypoglycemia in diabetics solely treated by diet [174]. If T2DM patients are treated with insulin they should take the same precautions as T1DM patients. Although frank hypoglycemia is rare, in patients treated with oral hypoglycemic agents, antidiabetic medication, extent of exercise, and food intake should, in principle, also be mutually adjusted [72]. This means that on exercise days the dosage of the drug may have to be reduced. Because weight loss usually is desirable in these patients carbohydrate supplement is less expedient.

### Table 41.1 General advice to the physically active insulin-dependent diabetes patient

- Be educated in exercise physiology and diabetes
- Use frequent insulin injections or pump therapy
- Reduce insulin dose prior to significant exercise
- Monitor blood glucose
- Take extra carbohydrate before, during, or after exercise
- Always carry readily digestible carbohydrate
- Have regular medical examinations
- Follow good exercise practice (e.g. regarding equipment, warm-up, technique, and hydration)
- Avoid extreme conditions or exercise with a healthy person
of 50–75% VO₂max is recommended, whereas high intensity with stable coronary artery disease dynamic aerobic activity accordingly. For hypertensive diabetics as well as for diabetics monitored initially, and adjustments in therapy should be made diabetics the blood pressure response to exercise should be particular, in hypertensive diabetics [183]. So, in hypertensive to exercise may be exaggerated in diabetics [179,182] and, in other hand, the increase in arterial blood pressure in response by mechanisms that are not well understood [179,181]. On the impaired during exercise in both type 1 and type 2 diabetics signs of cardiac impairment, left ventricular function may be impaired during exercise in both type 1 and type 2 diabetics by mechanisms that are not well understood [179,181]. On the other hand, the increase in arterial blood pressure in response to exercise may be exaggerated in diabetics [179,182] and, in particular, in hypertensive diabetics [183]. So, in hypertensive diabetics the blood pressure response to exercise should be monitored initially, and adjustments in therapy should be made accordingly. For hypertensive diabetics as well as for diabetics with stable coronary artery disease dynamic aerobic activity of 50–75% VO₂max is recommended, whereas high intensity activities known to elicit high increases in blood pressure, for example heavy resistance exercises involving breath-holding, should be minimized [179].

Autonomic neuropathy is associated with increased mortality from myocardial infarction and sudden death, and it may be a marker for clinically unrecognized cardiac disease [179]. Thus, silent ischemia upon exertion is more frequent in this condition because the anginal perceptual threshold may be increased in the presence of autonomic neuropathy [184,185]. Also diabetic cardiomyopathy with impaired left ventricular function at rest and during exercise is closely associated with cardiovascular autonomic neuropathy [186]. As is also prolonged QT interval in the ECG, and this predisposes to ventricular arrhythmias [187]. Because of the risk of an adverse cardiovascular event during exercise, a supervised graded exercise test is recommended prior to initiation of training programs in diabetics with autonomic neuropathy [131,187]. It should always be remembered that due to impaired heart rate response, in these patients rate of perceived exertion is a safer guide for exercise intensity than heart rate [187]. If the patients have orthostatic hypotension, which may be present in the absence of symptoms at rest, they may benefit from body stockings to increase venous return during exercise, and they will need long warm-up and cool-down periods of low-intensity exercise in relation to training sessions [179]. Another problem for patients with autonomic neuropathy may be reduced hypoglycemia awareness [187]. They may also have gastroparesis, which may delay the

Table 41.2: General guidelines for regulation of insulin-dependent diabetes in relation to exercise. Exact prescriptions cannot be given due to variation depending on the individual, the physical activity, and the environment

<table>
<thead>
<tr>
<th>Glucose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measure blood glucose before, during, and after exercise.</td>
</tr>
<tr>
<td>If blood glucose is higher than 14 mM (250 mg/100 mL) measure ketones in urine and do not exercise in case of ketonuria.</td>
</tr>
<tr>
<td>If blood glucose is higher than 16.5 mM (300 mg/100 mL) before exercise, be cautious even if no ketosis is present.</td>
</tr>
<tr>
<td>If blood glucose is below 5.5 mM (100 mg/100 mL) ingest extra carbohydrate, e.g. 20 g, before exercise.</td>
</tr>
<tr>
<td>Learn the glycemic response to various activity conditions.</td>
</tr>
<tr>
<td>Remember that hypoglycemia may develop after exercise, e.g. during the night following evening exercise.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Insulin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reduce dose of short-acting (maybe also of intermediate- or long-acting) insulin before exercise.</td>
</tr>
<tr>
<td>After exercise supplemental insulin may be necessary, e.g. if meals are increased. However, after extensive exercise insulin dose may have to be reduced.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Diet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do not exercise until 1–2 h after a main meal.</td>
</tr>
<tr>
<td>Take 15–25 g extra carbohydrate per 30 min exercise before unplanned postprandial exercise, when it is too late to reduce insulin.</td>
</tr>
<tr>
<td>Always carry readily digestible carbohydrate in case of hypoglycemia.</td>
</tr>
</tbody>
</table>

Table 41.3: The American Diabetes Association’s criteria for graded exercise testing with ECG and blood pressure measurement before increase in physical activity in diabetes patients

| Age >35 years |
| Age >25 years and type 2 diabetes for more than 10 years or type 1 diabetes for more than 15 years |
| Any additional risk factor for or overt coronary heart disease |
| Peripheral arteriosclerosis |
| Proliferative retinopathy or nephropathy including microalbuminuria |
| Autonomic neuropathy |

Source: Zinman 2003 [131].
intestinal absorption of ingested carbohydrate. So, more careful glucose monitoring is appropriate in relation to exercise. These patients may also have difficulties in thermoregulation and be prone to dehydration and electrolyte disturbances. For these reasons they are generally advised to avoid exercise in hot or cold environments and to be vigilant about adequate hydration [121,131,187]. Taken together, diabetics with autonomic neuropathy need extensive examination prior to embarking on an exercise program, and subsequent progression should be slow and carefully monitored.

Autonomic neuropathy may be accompanied by anhidrosis of the lower body resulting in dry, brittle skin that cracks easily. The tendency to fissures may be supported by compromised cutaneous vasomotion [187]. Also somatic neuropathy favors ulcer formation. Loss of protective sensation plays a major role allowing repeated trauma, the feet being in particular danger. However, motor neuropathy may also contribute by inducing interosseous muscle atrophy and, in turn, foot deformities [187,188]. Ulcers become easily infected due to poor antimicrobial defense mechanisms in diabetics, and their healing may also be impeded by peripheral atherosclerosis, which is frequently present in these patients. Leg muscle atrophy secondary to motor neuropathy as well as reduced proprioception and ensuing reduced reflex stabilization and coordination of movements predispose to traumatic injuries. In addition to ulceration such injuries may provoke a Charcot foot (neuroarthropathy) and, in turn, osteopenia predisposing to fractures of the small bones in the foot [187,189,190]. In order to avoid the potential, serious consequences of peripheral neuropathy, for example amputation, the diabetic patient suffering from this condition should inspect his or her feet daily for calluses, blisters, cuts, and scratches. Furthermore, the patient should maintain good foot hygiene, avoid wet feet, and use polyester or cotton socks and comfortable shoes with good pressure absorption. Repetitive weight-bearing exercises, for example prolonged walking and jogging, should be avoided [131,189]. Acceptable activities would be swimming, bicycling, and walking. Patients whose retinopathy has been adequately treated with laser therapy can probably undertake even strenuous exercise without undue risk after a period of 6 weeks.

In both type 1 and type 2 diabetics, exercise may provoke transient microalbuminuria, and if abnormal albumin excretion is present at rest exercise increases the albuminuria due to increased glomerular permeability [199,200]. Exercise-induced increases in albumin excretion correlate with exercise-induced increases in systolic blood pressure, and the latter may be exaggerated in diabetics [182,201]. The exercise-induced increase in albumin excretion can be ameliorated by both antihypertensive medical treatment and good metabolic control [200,202]. It is a matter of concern whether diabetics, who exercise regularly and thus have more exercise-induced albuminuria and blood pressure increases, run a risk of more rapid progression of nephropathy. However, comparisons between diabetics with different activity levels do not confirm this suspicion [203]. Although no permanent adverse effects have been documented, diabetics with microalbuminuria or overt albuminuria are nevertheless often dissuaded from strenuous physical activity [121,131,200]. Prior to embarking on a moderate- to high-intensity physical activity program these patients should be examined for macro- and microvascular disease including a graded exercise test with blood pressure monitoring, and those who suffer from hypertension should be well treated for this disease. In end-stage renal disease, probably reflecting chronic uremia, inactivity and accompanying autonomic neuropathy and heart disease, the aerobic exercise capacity is markedly reduced due to both reductions in oxygen carrying capacity of the blood (diminished maximal cardiac output and hemoglobin
concentration) and in oxygen extraction by muscle [204,205]. Also muscle strength is reduced [206,207]. The influence of uremia is illustrated by the fact that kidney transplantation leads to increases in hemoglobin and aerobic capacity [205]. Patients with end-stage renal disease can benefit from training, which may take place either on off-dialysis days or during hemodialysis [208]. The latter option offers a number of advantages, such as improved compliance to exercise, monitoring of potassium and glucose concentrations, and less muscle cramping and hypotensive episodes during dialysis. Patients treated by continuous ambulatory peritoneal dialysis have also been successfully trained [209]. Before initiation of an exercise program diabetics with end-stage renal disease should be screened for macro- and microvascular disease, the examination including a supervised graded exercise test [208]. Due to their susceptibility to fluid and electrolyte disturbances these patients are advised to avoid exercising in extreme environmental conditions [208].

Other conditions requiring special precautions
Type 2 diabetics are mostly overweight and, accordingly, more prone to overload injuries. Degenerative joint disease is more common in these patients. For these reasons, swimming, bicycling, rowing, and walking are preferable to pounding weight-bearing exercise such as jogging. During pregnancy the diabetic woman must abstain from more than mild exercise or even be prescribed bedrest if she develops medical complications such as hypertension or albuminuria or obstetrical complications, for example resulting in vaginal bleeding or evidence of intrauterine growth retardation [88,210]. Stroke and amputation are common disabling conditions in diabetic individuals. Rehabilitation requires therapeutic exercise, during which caution should be exerted due to the high risk of concurring diabetic long-term complications [211]. Age and duration of diabetes are independent criteria for supervised graded exercise testing of diabetics prior to training because these variables are related to the risk of unrecognized cardiovascular disease (Table 41.3) [131]. Finally, it has to be remembered that particularly older diabetics often take other drugs than insulin or oral hypoglycemic agents, and that these drugs may cause alterations in the metabolic and cardiovascular response to exercise and potentially limit physical performance [212]. For instance, β-adrenergic antagonists tend to reduce cardiac output and muscle blood flow and also promote development of hypoglycemia while masking the symptoms of this.

Guidelines for training in diabetes
For both type 1 and type 2 diabetics physical activity is accompanied by gains as well as risks. The physical hazards are generally higher and the metabolic profit lower in the type 1 than in the type 2 diabetic. On the other hand, the level of physical activity desirable for psychologic and social reasons is higher for type 1 diabetics because they are mostly younger.

Before increasing usual patterns of physical activity all diabetes patients should undergo a medical evaluation consisting of a careful medical history and a physical examination focusing on symptoms and signs of disease affecting the heart and blood vessels, eyes, kidneys, feet, and the autonomic and peripheral somatic nervous systems. If the diabetic is older than 35 years, or has had the disease for more than 10 years (type 2 diabetics) or 15 years (type 1 diabetics), or has complications or any additional risk factor for coronary artery disease, he or she should also be subjected to echocardiographic examination and a graded exercise test including ECG and blood pressure measurements (Table 41.3) [131]. Identification of areas of concern will allow appropriate treatment as well as individualized exercise prescription minimizing risk to the patient. As the disease is not stationary, the diabetic ought to see the doctor regularly.

Particularly for the type 1 diabetic patient, who wants to be physically active, it is of utmost importance that he or she—or the parents if the diabetic is a child—has a sound knowledge of the disease with a profound understanding of the interaction between food, insulin, and exercise in the regulation of the plasma glucose concentration (Table 41.1). The patients should be taught how healthy persons respond to exercise and how patients with diabetes differ from healthy individuals. Such knowledge will enable the diabetic to attempt to imitate the normal response by reducing the insulin dose prior to exercise and/or to take extra carbohydrate. However, the patients should also be aware of the fact that the effect of exercise on glucose homeostasis during and after exercise is not accurately predictable and may differ between and even within subjects [213,214]. This underlines the necessity of frequent blood glucose measurements as well as the need for readily available carbohydrate snacks. Furthermore, the diabetic should carry written information about him- or herself and his/her condition. It would appear that it is not possible to give exact recommendations on how to adjust insulin and diet for exercise that are broadly applicable [149]. The diabetic must also individually tailor his or her regimen based on experience. Combination of frequent blood glucose monitoring with multiple short-acting insulin injections or, even better, with insulin pump treatment, offers the highest flexibility in relation to engaging in unplanned exercise and the best glucose control. Rapidly evolving technology allows continuous monitoring of subcutaneous interstitial glucose concentrations and, by use of algorithms, prediction of ongoing development of blood glucose concentrations [215,216]. Proactive reactions can, accordingly, be initiated, and, in consequence, sound physical activity behaviors are increased [215,216]. Provided that the patients have been properly educated, intensified insulin therapy is not accompanied by a high risk of exercise-induced hypoglycemia [217,218]. Some rules are generally agreed upon [131]: If, before exercise, blood glucose is below 5.5 mM, then extra carbohydrate should be eaten before initiation of exercise. If, conversely, blood glucose is above 14 mM urine ketones should be measured, and if they are positive insulin should be taken.
and exercise delayed until ketones are negative. Even if ketones are negative, caution must be exerted if blood glucose is higher than 16.5 mM (Table 41.2).

The best glucose control is achieved if exercise is fit into a regular lifestyle in which insulin, food, and exercise are elements of a fixed schedule with the amount and timing of the components carefully adjusted. However, the more diabetics know about their disease, and the more familiar they become with their responses to exercise through blood glucose measurements, the greater a variation in physical activity from day to day they will be able to handle. From a medical point of view the recommended goal would be at least 30 min of exercise a day. For endurance type exercises moderate intensity corresponding to 50–75% of individual maximum aerobic capacity (%V\text{O}_{2}\text{max}, relative workload) is generally recommended. In subjects with a normal heart rate regulation, the relative workload in percent of maximum can be estimated as (actual heart rate minus resting heart rate) × 100/(maximum heart rate minus resting heart rate). If the individual maximum heart rate is not known, it can be roughly estimated as 220 minus age. Alternatively, perceived exertion may be evaluated from a scale ranging from 6 (very mild) to 20 (extremely heavy), and relative load be calculated as (actual level of exertion minus 6) × 100/14. It is of note that exercise does not have to be strenuous, even regular walking has beneficial effects, for example on cardiovascular health [166,219]. Much evidence indicates that the overall energy expenditure is more important than exercise intensity when it comes to health benefits [166,219–221]. Over recent years it has been recognized that resistance exercises can be beneficial, for example 2–3 times a week, in the training program of diabetics [126,156,222–225]. All the various skeletal muscle groups of the body may be engaged by this means. Just as for endurance exercises, the intensity of resistance exercises can be varied according to the fitness and health status of the diabetic patient, and responses to exercise can be monitored. Excessive blood pressure increases can be avoided by omitting any breath-holding. In both type 1 and type 2 diabetic patients strength training has been shown to elicit beneficial effects on insulin sensitivity and HbA1c levels similar to those seen with endurance exercise [127,148,151,160,226–228]. In addition, resistance training serves to maintain or improve muscle mass and strength, which is important, for example with aging [126,223,227,229,230]. Recently, short duration but high-intensity exercise has been advocated as a means to improve metabolic regulation in nondiabetics and also in patients with T2DM [231,232]. The advantage being that while this type of exercise induces metabolic adaptations similar to those that can be obtained through more traditional endurance training, it is time efficient and therefore offers a greater likelihood of beginning and maintaining an exercise program as lack of time is a commonly cited barrier to regular exercise participation.

From analysis of the diabetic patient’s condition and the demands of exercise, it is possible to predict the extent of a given physical activity that the patient will be able to tolerate. Diabetics in good metabolic control and without long-term complications can exercise on equal terms with healthy subjects. Adjustment of the therapeutic regimen will allow safe participation in all forms of physical activity even during changing daily schedules and preferences. In general, long-term changes in physical activity are difficult to achieve [233]. The fact that personal preferences can be complied with is important for maintenance of a physically active lifestyle. Activities that are considered to be enjoyable, varied, convenient in time and location, and which offer appreciable gains, are supervised and free, and can be performed in groups are likely to achieve the best compliance. In type 2 diabetics, motivational tools, for example pedometers, and an exercise consultation combining motivational theory and cognitive behavioral strategies also promotes an increase in physical activity [233–235]. Particularly in children, interactive computer games requiring vigorous body activity as input may enhance exercise behavior [236]. Appropriate exercise options for diabetics with long-term complications are suggested above and are described in more detail elsewhere [196]. For all kinds of training, in order to prevent musculoskeletal injuries and dropout for other reasons too, progression in exercise intensity and duration should be slow, and each training session should begin and end with muscle stretching and contractions of low intensity. Furthermore, optimum equipment (e.g. shoes) should be used. Exercise, which markedly increases cardiac output, should be avoided in the incubation and fever phases of infectious diseases, and caution should be exerted also in other acute or unstable chronic medical disorders. In case of unusual sensations (e.g. chest pain) exercise must be interrupted.

It is comforting, reassuring, and motivating for the diabetic to know that other patients meet the same troubles as him/her and that the disease does not necessarily restrain desired activities. In this context, the existence of summer camps, where diabetic children meet, play, and learn to handle their disease, should be remembered. Sharing of experiences among physically active people with diabetes may also take place through organizations like the Insulindependence organization (www.insulindependence.org) the mission of which is to enhance the quality of life for people with diabetes through exercise. Diabetics are involved in all kinds of sports, for example marathon running [237], competitive cross-country skiing [238], scuba diving [215,239], and mountain climbing [240], also at the professional level. There are examples from the past of diabetic athletes, who have become Olympic gold medalists and world champions [241,242], and growing knowledge and technology in diabetes care will enable this to become a more frequent event.

**Exercise in the prevention of diabetes**

Notwithstanding the therapeutic effect of exercise in patients who have developed diabetes, exercise may be particularly...
efficacious if instituted earlier in life. This supposition is in line with the fact that in experimental studies physical activity increases insulin sensitivity and improves glucose tolerance in both healthy subjects and in people with obesity, gestational diabetes, and impaired glucose tolerance as well as in first-degree relatives of patients with T2DM. More support for the ability of exercise to prevent T2DM has come from a multitude of descriptive studies of societies undergoing a transition from traditional lifestyles to more Westernized patterns, examination of migrants from traditional to more urbanized environments and vice versa, comparisons of active rural versus more sedentary urban populations, case-control studies, retrospective cohort studies, and prospective studies of active compared with sedentary individuals. The most powerful evidence has been provided by several nonrandomized or randomized, controlled interventional epidemiologic studies [1,156,243–247]. According to studies the risk of getting T2DM within a 6-year period can be reduced by at least 40% with an active compared with a sedentary lifestyle. Weight loss is not necessary for a beneficial effect of exercise, and an effect can be obtained by mild or moderate exercise, the effect depending on overall energy expenditure rather than on intensity of exercise [219]. While high-intensity low-volume exercise has recently been shown to improve metabolic status in type 2 diabetic patients [231,232] the long-term efficiency of this type of exercise in actual disease prevention or treatment has not yet been established. The most marked benefit of exercise in diabetes prevention has been observed in people with obesity or a family history of diabetes [244,248]. So, obvious target populations for training programs would be people at increased risk of T2DM, for example first-degree relatives, women with gestational diabetes, people with low birth weight or visceral obesity, and certain ethnic groups. However, translation of evidence from structured intensive trials to routine clinical settings is challenging [249].

References


CHAPTER 42

Sulfonylureas and meglitinides: insights into physiology and translational clinical utility

Harold E. Lebovitz1 and Arne Melander2

1 State University of New York Health Science Center at Brooklyn, Staten Island, NY, USA
2 Lund University, Malmö, Sweden

Key points
- Sulfonylureas and meglitinides are effective treatments for hyperglycemia in type 2 diabetes.
- Sulfonylureas and meglitinides bind to specific sites on pancreatic β-cell plasma membrane ATP-dependent potassium channels and close the channels by a glucose-independent mechanism. Closure of the ATP-dependent potassium channel results in membrane depolarization, opening of voltage-gated calcium channels, an increase in cytosolic calcium ion, and migration and exocytosis of insulin granules.
- Sulfonylureas and repaglinide at the appropriate doses improve glycemic control in type 2 diabetes similarly. Differences among the various agents are related to characteristics of: oral absorption, affinities and binding to SUR1 receptor sites, plasma half-life, mechanisms of metabolism, and sites of excretion.
- Side effects of glucose-independent insulin secretagogues are clinically significant hypoglycemia, weight gain, and possibly an increase in cardiovascular mortality.
- Sulfonylureas are unique treatments for diabetes caused by genetic abnormalities of glucose metabolism and ATP/ADP control of insulin secretion.

Historical perspective: the merger of clinical and basic science

There is no better example of the interplay of astute clinical observations leading to basic science research which then opens new insights into basic physiology and pathophysiology than the history of hypoglycemic sulfonylureas. As a consequence of food shortages and poor sanitation during World War II, epidemics of typhoid fever occurred in Europe. In 1942 Marcel Janbon, a clinician at the Clinic of the Montpellier Medical School in France had the occasion to treat such patients with a newly developed sulfonamide (2254 RP) [1]. He noted that many of the patients had adverse reactions such as convulsions and even prolonged comas. A major drop in blood glucose levels was observed in some patients. Dr. Janbon recognized the similarity of those findings to those of his patients taking insulin. He consulted with Professor August Loubatières of the physiology department who immediately undertook animal studies with 2254 RP. On June 13, 1942, Professor Loubatières observed that repeated oral administration of 2254 RP in a normal fasting conscious dog induced a progressive, severe and prolonged decrease in blood glucose. Thus he established the hypoglycemic effects of 2254 RP [1,2]. From 1942 through 1946 by undertaking a series of systematic studies of the effects of 2254 RP and structural analogues, he established that the hypoglycemic property of 2254 RP and its analogues was due to their ability to stimulate the secretion of insulin through a direct action on pancreatic islets. He proposed the use of hypoglycemic sulfonylureas to treat certain forms of diabetes [2].

For the next 50 years sulfonylureas were used to treat patients with T2DM without knowing the mechanism by which they were able to stimulate insulin secretion [3–6]. Additionally, a variety of in vivo and in vitro extrapancreatic effects of sulfonylureas were recognized, but neither the mechanism of these effects nor their relationships to the pancreatic effects were understood [7,8].

A series of in vitro studies in the 1980s established that sulfonylureas were able to stimulate insulin secretion without entering the β cell [7,9,10]. This led to the identification of a sulfonylurea receptor on the β-cell plasma membrane [11,12]. In 1995 the sulfonylurea receptor was cloned and the mechanism of sulfonylurea stimulation of insulin secretion was shown to be mediated by closure of the pancreatic β-cell ATP-dependent potassium (K\textsubscript{ATP}) channel [13,14]. Subsequent studies have identified multiple variants of the SUR receptor in tissues such as brain, myocardium, vascular endothelium and mitochondria.
explaining the mechanism whereby some sulfonylureas may exert extrapancreatic effects [15–17].

**Mechanism of action of sulfonylureas and meglitinides**

**Pancreatic β-cell effects**

ATP-dependent potassium (K$_{ATP}$) channels were discovered in the heart in 1983 [18]. Shortly thereafter, they were recognized in pancreatic β cells [19], brain, skeletal muscle, and many other tissues [15,16,20]. The cloning of the sulfonylurea receptor allowed the subsequent identification that the K$_{ATP}$ channel is a hetero-octamer comprising two subunits: four pore-forming Kir6.x subunits and four regulatory SUR (sulfonylurea receptor) subunits [14,20–22]. Kir6.x is a member of the inward rectifier K$^+$ channel family; SUR is a member of the ATP-binding cassette protein superfamily.

The K$_{ATP}$ channel functions to regulate potassium ion efflux from the cell and plays a role in maintaining the plasma membrane voltage potential. Closure of the cell membrane K$_{ATP}$ channel causes depolarization of the plasma membrane resulting in opening of plasma membrane voltage-gated calcium channels. Extracellular Ca$^{2+}$ enters the β cell increasing cytoplasmic Ca$^{2+}$ concentration (Figure 42.1). In secretory cells that increases granule exocytosis and hormone secretion (e.g. insulin release from β cells). In muscle cells there is an increase in contractility and in nerve cells there is an increase in signal transmission. The basal state in secretory cells is characterized by open K$_{ATP}$ channels which close during stimulation of secretory activity. By contrast, in tissues requiring high metabolic activity, the K$_{ATP}$ channels are ordinarily closed and open in response to the need to reduce energy requirements (hypoxemia) or to increase blood flow (ischemia) [16,20–22].

The state of the K$_{ATP}$ channel is determined by the regulatory subunits (SUR) which surround the rectifier subunits. The SUR subunit contains two nucleotide binding folds, three transmembrane domains and binding sites which recognize both the sulfonylurea moiety and the benzamido moiety (Figure 42.2) [20–22]. The nucleotide binding sites are located on the cytoplasmic surface of the K$_{ATP}$ channel and recognize the ATP/ADP concentration ratio. Low ATP/ADP ratios are associated with open K$_{ATP}$ channels. Increasing ATP/ADP ratios close increasing numbers of K$_{ATP}$ channels. The ATP/ADP concentration ratio is the physiologic regulator of the state of the K$_{ATP}$ channel. The sulfonylurea and benzamido binding sites of the SUR receptor are on the extracellular surface of the K$_{ATP}$ channel and their activation causes its closure independent of the ATP/ADP concentration ratio. Figure 42.2 is a schematic representation of the SUR 1 and Kir6.2 subunits of the pancreatic β-cell K$_{ATP}$ channel. Site A corresponds to the sulfonylurea binding site and Site B to the meglitinide (benzamido) binding site [21,22]. Drugs can be differentiated into those that bind to the sulfonylurea (A) site, those that bind to the benzamido (B) site, and those that bind to both sites as shown in Figure 42.3.

Oral hypoglycemic secretagogues which bind to either one or both sites cause closure of the pancreatic β-cell K$_{ATP}$ channel similar to but independent of the ATP/ADP concentration ratio. The result is depolarization of the plasma membrane and opening of the adjacent voltage-gated calcium channel (VDCC). Calcium ions from the extracellular compartment enter the β cell and increase the cytosolic Ca$^{2+}$ concentration [23]. The migration of the insulin granule to the plasma membrane and its exocytosis are proportional to the cytosolic Ca$^{2+}$ concentration. The potency and kinetics of the hypoglycemic insulin secretagogues are related to the binding site specificity and kinetics of interaction with the SUR1 subunit. Figure 42.4

![Figure 42.1](#) Scheme of a stimulated β cell. Sulfonylureas stimulate by binding to plasma membrane receptors that are coupled to ATP-dependent K$^+$ channels. This inhibits efflux of K$^+$, causing depolarization of the plasma membrane. As a consequence, voltage-gated Ca$^{2+}$ channels open, causing influx of Ca$^{2+}$ into the cytosol, which stimulates exocytosis of mature and immature insulin granules. ATP-dependent K$^+$ channels can also be inhibited by intracellular metabolic events that increase the ATP/ADP ratio. Source: Siconolfi-Baez et al. 1990 [12].
**Figure 42.2** Schematic of the Kir6.2 SUR1 subunit of the pancreatic β-cell ATP-dependent K⁺ channel. The SUR subunit contains three transmembrane domains. There are two nucleotide binding sites on the cytoplasmic surface (NBF-1 and NBF-2). Two binding sites are present (sulfonylurea and meglitinide) which can interact with sulfonylurea and related molecules. The channel can close when ATP/ADP ratios are high or independently of ATP/ADP ratio when activated by molecules which bind to either the sulfonylurea or meglitinide sites. Source: Seino 2012 [22]. Reproduced with permission of John Wiley & Sons.

**Figure 42.3** Molecules which bind to either the sulfonylurea (A) site, the meglitinide (B) site or both sites. Binding to either or both sites closes the ATP-dependent channel and stimulates insulin secretion. Source: Modified from Melander et al. [6]. Reproduced with permission of Springer.
Chapter 42

Glipizide
Glibenclamide
Tolbutamide
Chlorpropamide

Figure 42.4 Displacement of [3H]glipizide from purified plasma membranes from RINm5F-generated insulinomas by glipizide (●), glyburide (x), tolbimate (▲) and chlorpropamide (○). Data are mean ± SE of three determinations. Source: Siconolfi-Baez et al. 1990 [12].

depicts the binding site specificity of the more commonly used hypoglycemic insulin secretagogues.

There are several additional sulfonylurea binding sites in the β cell, the significance of which are controversial. Studies have shown that 75% of glibenclamide binding to the β cell is to the insulin granule [24]. This binding has been postulated to activate processes which facilitate granule exocytosis. More recently sulfonylureas have been postulated to bind to and activate Epac2, an exchange protein directly activated by cyclic AMP [25]. This protein acts as a guanine nucleotide exchanger for Rap-1 and promotes insulin granule exocytosis [26]. Epac2 is known to play a significant role in cyclic AMP-mediated insulin secretion [26]. Sulfonylurea binding and activation of Epac2 has been controversial with some positive and some negative data [25,27].

The SUR subunit of the K\textsubscript{ATP} channel of the myocardial (SUR2A) and vascular smooth muscle cells (SUR2B) are isoforms of SUR1. SUR2A has 68% homology with SUR1 and SUR2B differs from SUR2A by 42 amino acid residues in the C-terminus [15,16]. Their amino acid sequences differ somewhat from that of the SUR1 subunit in the K\textsubscript{ATP} channels of β cells and neurons (SUR1). These differences include the binding site regions for sulfonylureas and meglitinides and provide for individual sulfonylurea-mediated differences in extrapancreatic effects [15,16]. Extrapancreatic effects of sulfonylureas are dependent on the specific SUR subunit of the tissues K\textsubscript{ATP} channel, the concentration of the sulfonylurea achieved at the tissue’s K\textsubscript{ATP} channel and the affinity of the sulfonylurea for the specific SUR subunit in that tissue’s K\textsubscript{ATP} channel [28–32]. In vitro demonstration of an extrapancreatic effect of a sulfonylurea may have no physiologic or pharmacologic equivalent in vivo since tissue concentration achieved in vitro is many times greater than that achievable in vivo and accessibility to the tissue is not limited in vitro as it is in vivo.

**Extrapancreatic effects of sulfonylureas**

K\textsubscript{ATP} channels are present in many cells and regulate a variety of physiologic processes (Table 42.1). The Kir0 subunit has at least two isoforms and the SUR subunit has three isoforms. Combining different subunit isoforms creates K\textsubscript{ATP} channels with differing characteristics of the binding sites to serve the physiology of the different tissues. The K\textsubscript{ATP} channel of the pancreatic β cell is Kir6.2 SUR1, that of myocardial (sarclemma) and skeletal muscle cells is Kir6.2 SUR2A, that of smooth muscle cells Kir6.2 SUR2B, that of vascular smooth muscle Kir6.1 SUR2B, that of neurons Kir6.2 SUR1, and mitochondrial K\textsubscript{ATP} channels are closely mimicked by Kir6.1 SUR1 subtype [22].

The K\textsubscript{ATP} channels in all tissues regulate the extrusion of potassium ions from the intracellular compartment and the flow of extracellular calcium into the intracellular cytosol. Depending on the basal state and physiology of the tissue, these changes influence neuronal activity in the hypothalamus which responds to changes in nutrients (hypoglycemia) or stimuli coming from other neuronal centers (ischemia) [30,31], energy metabolism, contractile force and arrhythmias in the myocardium [32], blood flow in arteries [22], and contractile force in skeletal muscles [22]. The ability of a particular sulfonylurea or meglitinide to affect these activities will depend on the specificity and binding characteristics of the drug for the SUR binding site of the tissue. The concentration of the drug achieved at the binding site may be equally as important as, for example, neuronal activation where the SUR subunit is the same as in the pancreatic β cell.

**Pharmacology and clinical pharmacology**

**Chemistry and structure of sulfonylureas and meglitinides**

The sulfonylureas are sulfonamide derivatives (Figure 42.3). For years it was presumed that the hypoglycemic effects of the sulfonylureas required the sulfonylurea moiety. The demonstrations in the 1980s that the nonsulfonylurea analogues of glibenclamide and gliclizide mimic the in vitro and in vivo effects of their parent compounds on pancreatic β-cell function and insulin release indicated that the structural requirements
for stimulating insulin secretion did not necessarily require the sulfonylurea moiety [33]. This finding has subsequently been explained by the demonstration of the two discrete binding sites on the SUR1 receptor [21,22]. It was also the basis for the development of the meglitinides (repaglinide) [34]. The sulfonylureas have been described as “first generation sulfonylureas” and “second generation sulfonylureas.” First generation sulfonylureas, such as tolbutamide and chlorpropamide, have a phenyl ring with simple substituents at one end and an aliphatic side chain at the other end of the molecule (Figure 42.3). In the second generation sulfonylureas, such as glibenclamide (glyburide), glipizide, and glimepiride, the aliphatic side chain has been replaced by a cyclohexyl group and the substituent at the other end is a more complex structure than in the first generation agents. The substitutions markedly increase the molecular affinity of the second generation agents to the SUR1 subunit of the β-cell K\(_{\text{ATP}}\) channel as compared to the first generation agents and explain their higher intrinsic activity and potency (Figure 42.4).

**Pancreatic β-cell K\(_{\text{ATP}}\) channel and insulin secretion**

The pancreatic β cell has multiple modes by which it can increase insulin secretion (Figure 42.5) [21,26,35]. The primary mode is that involved in controlling glucose-mediated insulin secretion and it is regulated by the plasma membrane K\(_{\text{ATP}}\) channel [26,35]. The ambient plasma glucose concentration determines the pancreatic β-cell ATP/ADP ratio. Glucose entry into the β cell is controlled by the specific glucose transporter, GLUT2, which has high capacity and low affinity for glucose and allows β-cell glucose levels to rapidly equilibrate with plasma glucose levels. Pancreatic glucokinase quickly converts the glucose to glucose-6 phosphate which is oxidized by the mitochondria to generate increased ATP and decreased ADP levels. The intracellular ATP/ADP ratio is the pancreatic β-cells glucose sensor.

When the plasma glucose is low, the ATP/ADP ratio is low and the K\(_{\text{ATP}}\) channel is open, extruding potassium ions from the β cell, and is maintaining the cell membrane voltage potential (Figure 42.6). The adjacent voltage-gated calcium channel is closed and insulin secretion is low. As the plasma glucose increases, the ATP/ADP ratio rises and K\(_{\text{ATP}}\) channels close (Figure 42.6). The plasma membrane depolarizes, opening the plasma membrane voltage-gated calcium channels. Extracellular calcium enters the cell and increases cytosolic calcium ion concentrations. The increased calcium ion concentration stimulates insulin granule movement to the plasma membrane and its release into the circulation.

Glucose via ATP/ADP stimulates insulin secretion from the cytosolic surface of the K\(_{\text{ATP}}\) channel. Sulfonylureas and meglitinides bind to the SUR1 receptor on the extracellular surface of the K\(_{\text{ATP}}\) channel and stimulate insulin secretion independent of the ambient plasma glucose level. These agents are glucose-independent insulin secretagogues.

**Acute pancreatic effects of sulfonylureas**

Initial studies on the hypoglycemic effects of sulfonylureas demonstrated that functional pancreatic islets were necessary for this action [2]. Sulfonylurea administration has no effect on blood glucose in pancreatectomized or alloxan diabetic animals [2] or in humans with established type 1 [36] or pancreatic diabetes [37]. Acute intravenous or oral administration of sulfonylurea leads to a rapid rise in portal and then peripheral insulin and C-peptide levels in normal or type 2 diabetic humans or animals [38]. Direct stimulation of insulin release by sulfonylureas has been demonstrated in vitro utilizing perfused pancreas preparations, isolated pancreatic islets, and β-cell cultures [4,39]. Sulfonylurea stimulation of insulin secretion

### Table 42.1 Tissue specificity and physiologic role of K\(_{\text{ATP}}\) channels in different tissues

<table>
<thead>
<tr>
<th>Tissue</th>
<th>Rectifier subunit</th>
<th>SUR subunit</th>
<th>Basal state of K(_{\text{ATP}}) channel</th>
<th>Stimulus</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pancreatic β-cell</td>
<td>Kir6.2</td>
<td>SUR1</td>
<td>Open</td>
<td>Glucose closes</td>
<td>Insulin secretion</td>
</tr>
<tr>
<td>Brain</td>
<td>Kir6.2</td>
<td>SUR1</td>
<td>Dependent on ATP/ADP</td>
<td>Hypoxia, hypoglycemia, nerve injury, and nutrient flux open</td>
<td>Counterregulation from hypoglycemia, reduces membrane excitability, protects from seizures, nutrient regulation, increases K(^{+}) efflux, decreases Ca(^{2+}) entry, reduces energy consumption, increases re-entrant arrhythmias</td>
</tr>
<tr>
<td>Cardiac sarcolemma</td>
<td>Kir6.2</td>
<td>SUR2A</td>
<td>Closed</td>
<td>Hypoxia and ischemia open</td>
<td></td>
</tr>
<tr>
<td>Cardiac mitochondria</td>
<td>Kir6.1</td>
<td>SUR1</td>
<td>Closed</td>
<td>Hypoxia opens</td>
<td>Muscle activity opens</td>
</tr>
<tr>
<td>Skeletal muscle</td>
<td>Kir6.2</td>
<td>SUR2A</td>
<td>Closed</td>
<td>Hypoxia opens</td>
<td>Muscle activity opens</td>
</tr>
<tr>
<td>Vascular smooth muscle</td>
<td>Kir6.1</td>
<td>SUR2B</td>
<td>Closed</td>
<td>Hypoxia and ischemia open</td>
<td>Vasodilatation</td>
</tr>
<tr>
<td>Smooth muscle</td>
<td>Kir6.2</td>
<td>SUR2B</td>
<td>Closed</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Notes

1. **ATP** is adenosine triphosphate, a high-energy compound that provides energy for cellular processes.
2. **ADP** is adenosine diphosphate, a compound involved in energy transfer reactions.
3. **K\(_{\text{ATP}}\)** is the potassium channel regulated by ATP/ADP ratios.
4. **Glucokinase** is an enzyme that converts glucose to glucose-6 phosphate.
5. **GLUT2** is a glucose transporter that facilitates the movement of glucose across the cell membrane.
6. **C-peptide** is a hormone that helps regulate insulin secretion.
7. **K\(^{+}\)** is potassium ions.
8. **Ca\(^{2+}\)** is calcium ions.
9. **Sulfonylureas** are a class of drugs used to treat diabetes by stimulating insulin release.
10. **Meglitinides** are another class of drugs used to treat diabetes, similar to sulfonylureas but with different mechanisms of action.

---

**Figure 42.3** shows the molecular structure of sulfonylureas, with the first generation agents having a phenyl ring and the second generation agents having a cyclohexyl group. **Figure 42.4** illustrates the regulation of K\(_{\text{ATP}}\) channels in different tissues, highlighting the importance of glucose concentration in determining insulin secretion. **Figure 42.5** and **Figure 42.6** depict the acute effects of glucose and ATP/ADP on insulin secretion in pancreatic β-cells.
Multiple modes of insulin secretion. Glucose stimulation of insulin secretion is mediated by the ATP/ADP ratio regulation of the ATP-dependent $K^+$ channel. Extracellular glucose is transported into the $\beta$ cell by the GLUT2 transporter, is phosphorylated by glucokinase and generates ATP via the electron transport chain. High ATP/ADP ratios cause closure of the channel. Sulfonylureas and meglitides cause closure of the ATP-dependent $K^+$ channel independent of glucose regulation of ATP/ADP ratio. Incretins bind to G-protein regulated cAMP-dependent receptors. Their activation increases protein kinase A-regulated pathways and amplifies insulin secretion in the presence of elevated plasma glucose levels. Granule migration is facilitated by Epac2 which is activated by cAMP.

Coordinated regulation of ATP-dependent $K^+$ and voltage-gated $Ca^{2+}$ channels in various tissues. Low ATP/ADP levels are associated with low glucose ($\beta$ cells), low energy requirements (heart), hypoxia (heart), and ischemia (heart). Such states are characterized by high potassium extrusion, maintenance of membrane potential, and low intracellular $Ca^{2+}$. High ATP/ADP levels are associated with hyperglycemia ($\beta$ cells), increased energy requirements, normal heart, normal oxygenation and normal blood flow. Potassium extrusion is minimized, $Ca^{2+}$ entry into the cytoplasm is maximized, action potentials are normal and energy expenditure is high.

from normal $\beta$ cells is usually weaker than nutrient stimulation, but the reverse is frequently true with diabetic $\beta$ cells.

A number of observations had suggested that the initial event in sulfonylurea stimulation of insulin secretion involved binding of the sulfonylurea molecule to a specific site on the $\beta$-cell plasma membrane. Early studies of the distribution of tolbutamide in vivo concluded that the drug was restricted to the extracellular compartment with the possible exception of the liver [7]. The uptake of $[^{35}S]$tolbutamide into microdissected pancreatic islets was found to be identical to that of markers known to equilibrate with the extracellular space [4]. When perfused through isolated rat pancreas, dextran-linked tolbutamide was shown to be as effective as tolbutamide in stimulating first-phase insulin release [10].

Pursuing the concept that sulfonylurea action involves drug interaction with the plasma membrane, several groups of investigators demonstrated that radiolabeled sulfonylureas such as $[^3H]$glipizide or $[^3H]$glibenclamide bind to purified plasma membranes from either rat insulinoma or transformed hamster tumor $\beta$ cells with characteristics that are compatible with specific receptor binding [11,12]. Displacement by sulfonylureas occurs in proportion to their activity (Figure 42.4). Patch clamp techniques have shown a potency order of glibenclamide > glipizide > tolbutamide in the inhibition of the ATP-dependent potassium current in insulin-secreting cells [11].

Within several years, the sulfonylurea binding site was isolated, cloned and identified as the SUR subunit of the $\beta$-cell plasma membrane $K_{\text{ATP}}$ channel [13,14]. Sulfonylurea binding to the $K_{\text{ATP}}$ channel initiates events similar to those occurring with glucose stimulation (Figure 42.1). Sulfonylureas inhibit $^{86}$Rb efflux (used as a tracer for K+ movement) from islets
incubated at basal glucose concentrations [39]. The stimulatory effect of sulfonylureas on insulin secretion is abolished in Ca^{2+}-deprived media [23], and blockage of Ca^{2+} entry by agents such as verapamil causes a dose-dependent inhibition of sulfonylurea-mediated insulin secretion [23]. Sulfonylureas cause a two- to threefold increase in β-cell cytosolic Ca^{2+} as determined by quin-2 fluorescence [23,39]. Thus, the ion flux changes seen with sulfonylurea stimulation of β cells are similar to that following glucose stimulation. In both instances the rise in cytosolic Ca^{2+} activates the microtubular-microfilamentous system in β cells through one or more of several proposed mechanisms involving protein kinase phosphorylation [39].

While the delineation of the interaction between sulfonylureas and the β-cell plasma membrane K_{ATP} channel explains much of the mechanism by which sulfonylureas stimulate insulin secretion, a number of questions and observations remain unanswered. Several studies have demonstrated that as much as 75% of glibenclamide that attaches to the β cell is actually bound to β-cell granules [24]. These data are consistent with an observed direct effect of sulfonylureas in stimulating the exocytotic process itself. Some studies have suggested that sulfonylureas bind to a 65-kDa receptor in the β-cell granule membrane and activate granular CIC-3 chloride channels which are postulated to promote acidification of the granules which then gain release competence. As previously noted, Zhang and coworkers have recently described a direct effect of sulfonylureas in activating Epac2 which is known to be involved in cyclic AMP-mediated β-cell granule migration and release [25].

The K_{ATP} channel model does not explain the known differences between nutrient-stimulated insulin secretion and sulfonylurea stimulation of insulin secretion. D-Glucose is thought to stimulate insulin secretion by increasing ATP and/or modifying other factors that close the K_{ATP} channel from the cytoplasmic surface of the membrane [39]. Sulfonylureas differ from D-glucose in that they do not increase intracellular ATP or increase the efflux of radioactive phosphate from prelabeled islets [39]. In contrast to glucose, sulfonylureas decrease K^+ inflow into islet cells, increase their Na^+ content, and lower their intracellular pH [39]. Similarly, they do not stimulate proinsulin biosynthesis, neither do they inhibit [2-^{3}H]-adenosine efflux from prelabeled islets. Sulfonylureas only stimulate first-phase insulin secretion [39]. Thus there are many differences between sulfonylurea-mediated and nutrient-mediated insulin secretion.

The presence of sulfonylurea and meglitinide binding sites on the SUR subunit raises the question of whether there are natural ligands for these binding sites. If so, what are they and are alterations in the ligands or in the ATP-dependent potassium channel playing a role in the pathogenesis of β-cell insufficiency in T2DM.

**Chronic effects of sulfonylureas**
Sulfonylureas have been administered intravenously or orally to humans. Intravenous administration of 1 g tolbutamide bolus has been used as a test to assess insulin and/or C-peptide secretion [40]. The release of insulin occurs within minutes and is usually complete within 1 hour.

Orally administered sulfonylureas have been used for the chronic treatment of T2DM [3,5,6,22,41]. All sulfonylureas have equivalent effects on insulin release and blood glucose lowering when given at their maximal effective dose. The dose that gives the maximal effect of a given sulfonylurea depends on the affinity and binding characteristics of the specific sulfonylurea for the β-cell SUR1 subunit and the pharmacokinetic characteristics and metabolism of the drug. Sulfonylureas can be clinically classified by their duration of action and their mechanism of metabolism and excretion.

In numerous clinical studies it was shown that chronic treatment with sulfonylureas, in contrast to acute treatment, resulted in improved glycemia but with less rather than more insulin secretion [3,8]. Similarly when patients on chronic sulfonylurea therapy showed deteriorating glycemic control it was assumed that the sulfonylurea drugs had lost their insulin secretory activity. Assessments of the chronic effects of sulfonylureas on insulin secretion in patients with T2DM have been difficult to interpret because the magnitude of sulfonylurea stimulation of insulin secretion is influenced by many factors. Some of the factors that influence insulin secretion during chronic sulfonylurea therapy are (a) the intrinsic effect of sulfonylureas on β-cell insulin secretion; (b) the level of glycemia attained during the sulfonylurea stimulus for insulin secretion; (c) the degree of chronic fasting hyperglycemia, which itself may cause inhibition of insulin secretion (glucose toxicity); (d) the stage of glucose intolerance (the functional state of the β cells); and (e) tachyphylaxis to sulfonylureas during chronic, continuous sulfonylurea therapy.

Many studies in type 2 diabetic patients indicate that chronic sulfonylurea treatment increases the ability of the β cell to respond to a fixed stimulus. Porte and colleagues have defined the insulin secretory responses of nonglucose stimuli at different ambient glycemic levels in normal subjects and in patients with T2DM [42]. Normal individuals have an increase in insulin secretion stimulated by intravenous arginine or isoproterenol during an acute intravenous tolbutamide infusion, provided that the plasma glucose level is clamped, but not if it is allowed to fall [42]. It was also shown that 4 of 9 type 2 diabetic patients treated with chlorpropamide for 12–16 weeks had an increase in nonglucose-mediated insulin secretion when their pre-treatment glycemic values were restored [43]. The same phenomenon may explain why type 2 diabetic patients on chronic sulfonylurea therapy will show a rise in nutrient-mediated insulin responses immediately following discontinuation of sulfonylurea therapy [44]. A chronic effect of sulfonylurea therapy to increase the insulin secretory capacity of laboratory animal pancreas to stimuli has not been observed; rather, nutrient-stimulated insulin secretion is decreased.

Studies in type 2 diabetic patients with normal or mildly elevated fasting plasma glucose concentrations show that the
degree of the nutrient-evoked rise in plasma glucose is a major determinant of insulin secretion [5]: thus, a lesser rise such as that occurring during chronic sulfonylurea therapy will result in less secretion.

The chronic level of fasting hyperglycemia is as important as the nutrient-evoked rise in plasma glucose in determining insulin secretion. In this case, however, the higher the fasting plasma glucose value, the lower the nutrient-mediated insulin secretion. A reduction in fasting hyperglycemia by any form of therapy (diet, drugs, or insulin) results in increased glucose-mediated insulin secretion [45]. Exhaustive experimental and clinical data support the concept that chronic hyperglycemia itself inhibits glucose-mediated insulin secretion [46]. Apparently, reduction of fasting and postprandial hyperglycemia in type 2 diabetic patients by any modality can either increase or decrease glucose-mediated insulin secretion.

The question of sulfonylurea tachyphylaxis has been raised by Karam et al. [47], who studied the acute insulin secretory response to intravenous tolbutamide or glucagon in 10 type 2 diabetic patients, before, during, and after tolazamide therapy. They found that chronic tolazamide therapy abolished the insulin secretory response to tolbutamide but not to glucagon, and they also showed that the acute response to tolbutamide reappeared after cessation of tolazamide therapy. They concluded that chronic continuous sulfonylurea therapy may blunt the acute insulin secretory response to sulfonylurea. In vitro studies have shown that prolonged incubation of β cells with sulfonylureas causes a disturbance in the assembly and functional activity of the ATP-dependent potassium channel [48]. It is unclear whether this mechanism is operative in humans. Another possibility is that this phenomenon could be explained by downregulation of the SUR1 receptor. Support for this assumption can be obtained from two studies discussed subsequently which showed that increases of dosage and sulfonylurea steady-state concentrations over a certain level impaired rather than improved glucose and insulin responses.

This complexity explains why so much controversy exists as to whether the primary action of chronic sulfonylurea therapy is increased insulin secretion and why there has been continuing interest in whether extrapancreatic effects of sulfonylureas are clinically meaningful.

Effects on proinsulin biosynthesis

Normal α cells do not appear to have functional ATP-sensitive potassium channels and do not show significant in vitro effects of sulfonylureas in regulating glucagon secretion. However, several studies in humans show that glibenclamide treatment causes impairment of glucagon secretion in response to insulin-induced hypoglycemia [54–56]. The mechanism whereby sulfonylureas reduce glucagon secretion is thought to be an increase in intra-islet insulin flow from β cells to α cells. β-Cells contain K\textsubscript{ATP} channels with SUR subunits. Sulfonylureas stimulate pancreatic β-cell somatostatin release [54,56] but it is not clear whether this plays any part in their antidiabetic actions.

Extrapancreatic effects of sulfonylureas

Sulfonylurea drugs have been investigated extensively for extrapancreatic actions [8,57]. Most studies demonstrating such actions have utilized organ perfusion, cell culture, or subcellular preparations from laboratory animals. Some have shown extrapancreatic effects with concentrations of sulfonylureas far in excess of ordinarily attainable therapeutic plasma levels. Studies in vivo have frequently failed to validate that the action is a primary one rather than a consequence of better glycemic regulation due to increased insulin secretion. The two sulfonylureas that have been promoted most as having clinically significant extrapancreatic effects in humans are gliclazide [22] and glimepiride [58,59].

As sulfonylureas lower blood glucose only in the presence of a functioning pancreas, it is difficult to attribute a meaningful antidiabetic action to an extrapancreatic effect that occurs in the absence of insulin [59]. Thus, a primary action in increasing fructose 2,6-bisphosphate in the liver with subsequent stimulation of glycolysis and inhibition of gluconeogenesis, although of interest, is unlikely to be a significant component of sulfonylurea antidiabetic action. A similar analysis can be made for primary actions on muscle or adipose tissue.
**Effects on insulin action**

An extrapancreatic action of sulfonylureas that may have a meaningful antidiabetic effect is the potentiation of insulin action. Such an action was described in 1955 and has been controversial ever since. The earliest demonstrations of sulfonylurea potentiation of insulin action have been carried out in animal models. Experiments utilizing *in vitro* mouse diaphragms or organ cultures with rat adipose tissue have demonstrated that ordinary doses or concentrations of sulfonylureas potentiate insulin-mediated 2-deoxy-D-glucose uptake following a latent period of about 12 h [60]. One study suggested that this is attributable to sulfonylurea potentiation of insulin-mediated translocation of the glucose transporter from intracellular storage sites to the plasma membrane [61].

Putnam et al. [62] utilized the euglycemic insulin clamp at physiologic insulin levels in normal dogs and showed that 2 weeks’ treatment with glipizide caused a twofold increase in insulin-mediated glucose disposal. The chronic treatment caused no significant alterations in plasma glucose or insulin.

Studies in patients with T2DM that have attempted to demonstrate sulfonylureas potentiation of insulin-mediated glucose disposal have reported diverse results. Some show that sulfonylurea treatment increased insulin-mediated glucose disposal, but only in patients who have increased insulin secretion [63]; the improved glucose disposal is attributed to higher plasma insulin levels. In a study on insulin-treated type 2 diabetic subjects who had little or no remaining endogenous insulin, sulfonylurea failed to evoke any glucose reduction [64]. Other studies show that sulfonylurea treatment increases insulin-mediated glucose disposal, but the improvement was attributed to the effect of better glycemic regulation. Finally, there are studies in which sulfonylurea treatment improves glycemic control, increases insulin-mediated glucose disposal, and the effects do not correlate with improved insulin secretion, suggesting that true sulfonylurea-induced potentiation of insulin action does occur in some patients with T2DM [65,66].

The liver may be a site of sulfonylurea potentiation of insulin action. Sulfonylureas have been reported to potentiate insulin-mediated glycogen and lipid synthesis. These effects have been described in hepatic cell cultures. Clinical studies suggest that the reduction in fasting hyperglycemia during sulfonylurea treatment can be attributed in part to potentiation of insulin action on glycogen synthesis and gluconeogenesis [64]. Detailed discussions of the relative roles of pancreatic versus extrapancreatic mechanisms in the antidiabetic action of sulfonylureas are available in several reviews [57,67].

The mechanisms of the various extrapancreatic actions of sulfonylureas have not been defined. ATP-dependent potassium channels and specific plasma membrane sulfonylurea binding sites have been found in many tissues, as noted in Table 42.1. They have not been found in liver cells and are present in very minimal quantities in skeletal muscle cells. There is no evidence to suggest that ionic fluxes have any role in mediating the extrapancreatic effects described. The insulin-potentiating action of sulfonylureas has a latent period of several hours, and this effect is blocked by cycloheximide, indicating that new protein synthesis may be necessary for that effect to occur [68].

The effect of sulfonylureas in potentiating insulin action on muscle, adipose tissue, and liver has raised the issue of whether these effects are mediated by a direct action in increasing insulin binding by the insulin receptor (an increase either in receptor number or in affinity) or through post-receptor mechanisms. Studies *in vivo* in which insulin binding is assessed are invalid to answer this question because changes in plasma insulin levels will influence the number of insulin receptors, and measurements of insulin binding to circulating monocytes or adipose tissues may not reflect what is occurring at liver and muscle cells. Assessment of insulin binding to fibroblasts in culture has little physiologic meaning. Several studies *in vitro* indicate that the effect occurs at a post-receptor site [57].

**Effect on systemic availability of insulin**

One of the earliest extrapancreatic actions of sulfonylureas to be described was inhibition of insulinase, particularly in the liver [69]. Later studies disputed these data, and the idea was abandoned. However, several recent studies suggest that at least glipizide [70,71] and glibenclamide [72] increase the systemic availability of insulin through reduction of the hepatic extraction of insulin secreted from the pancreas. It is not known whether the effect is a primary one, for example inhibition of hepatic insulinase or displacement of insulin from hepatic binding sites, or is secondary to the increased rate of insulin secretion subsequent to sulfonylurea stimulation of the β cells [71]. If the increased systemic availability were attributable to displacement of insulin from hepatic insulin receptors, it is also possible that such displacement may reduce the effect of insulin exerted on the liver and hence would lead to an increased hepatic output of glucose, provided that the displacement was extensive enough. Such a phenomenon would help to explain why high doses of sulfonylurea may impair instead of improve blood glucose control [73,74].

**Effects on blood lipids**

Sulfonylurea treatment has been reported to have beneficial, neutral, or adverse effects on blood lipids [5,6]. It seems unlikely, therefore, that sulfonylureas have direct effects on very low-density lipid triglycerides, low-density lipid cholesterol, or high-density lipid cholesterol.

**Effect on fibrinolysis**

A recent study on bovine aortic endothelial cells indicated that certain sulfonylureas may enhance fibrinolysis and that the fibrinolytic mechanism is associated with sulfonylurea-enhanced production of plasminogen activator [75].

**Differences in extrapancreatic effects**

Some extrapancreatic effects of sulfonylureas are related to parts of the molecule other than the sulfonylurea core and are
therefore unrelated to the antidiabetic action. These extrapancreatic effects may confer unique advantages or disadvantages to a particular sulfonylurea. Chlorpropamide facilitates both the release and the tubular action of vasopressin, causing water retention and hyponatremia in certain patients [76]. Sulfonylureas increase the synthesis and secretion of plasminogen activator from bovine aortic endothelial cells, but the potency is unrelated to the blood glucose lowering activity [75]. The spectrum of action of glibenclamide on the myocardium differs from that of tolbutamide and glimepiride, as discussed in a later section. The clinical potential of many of these differences has yet to be explored.

**Clinical aspects of sulfonylurea therapy**

Knowledge about the clinical effects of sulfonylureas can be divided into that obtained prior to and after 1995. Studies from the 1950s through 1995 were primarily observational with little of the rigorous design and implementation of current studies. Endpoints were primarily changes in fasting blood glucose levels with values <100 mg dL\(^{-1}\) being defined as excellent, between 100 and 130 mg dL\(^{-1}\) as good, and >130 mg dL\(^{-1}\) unsatisfactory [77]. Those early studies can be summarized as follows. Sulfonylureas were effective in improving glycemic control in patients with T2DM with 60–65% achieving excellent or good control [77–79]. Sulfonylureas increased insulin secretion by a glucose-independent mechanism. Sulfonylureas had their primary effect in lowering fasting plasma glucose through decreasing hepatic glucose production [80–82]. Sulfonylureas had only a slight effect in decreasing postprandial glucose excursions because they did not significantly improve the delay in meal-mediated insulin secretion which is characteristic of T2DM [83]. They significantly increased late meal-mediated insulin secretion which frequently led to hypoglycemia [82,83]. Twenty to 25% of patients initially had little or no improvement in glycemic control (primary failures) [79,84,85]. An initially successful therapeutic response to sulfonylureas was often followed by secondary failure which occurred at a rate of 5–10% per year [79,84,85]. The secondary failure rate was similar for all of the first generation sulfonylureas [86].

Since 1995, studies have utilized changes in HbA1c as the primary efficacy of sulfonylurea therapy. It is important in evaluating such studies to recognize that for all therapeutic treatments, the decrease in HbA1c is a function of the starting (baseline) HbA1c [87]. That is, the higher the starting HbA1c the greater the decline and the lower the starting HbA1c the less will be the decline. This observation suggests that the therapeutic response to an antihyperglycemic agent is a composite of the agent’s mechanism of action and the improvement in glucose toxicity that occurs from the treatment. The importance of this observation is that therapeutic agents can only be compared in the same, randomized, controlled clinical trial. When sulfonylureas are used as primary therapy added to patients inadequately controlled by lifestyle modification, the mean decrease in HbA1c is 1.5% at a mean baseline HbA1c of 8.5% [65,81,82]. When added as a second or third agent for treatment, the incremental decrease in HbA1c is likely to be 0.6–0.8% [41,82,88].

Much of our long-term information about sulfonylurea therapy for patients with T2DM comes from the United Kingdom Prospective Diabetes Study (UKPDS) which was started in the late 1970s and concluded in 1997. This study recruited slightly more than 5000 newly diagnosed type 2 diabetic patients and after a 3-month diet intervention, randomized those who were still hyperglycemic to either continued lifestyle modification without drug addition (“conventional treatment”) or continued lifestyle modification to which was added a sulfonylurea (glibenclamide, chlorpropamide, or glipizide) or insulin (“intensive control”) [89,90]. Some overweight patients were randomized to addition or no addition of metformin, and in a substudy, some normal-weight and overweight patients already on sulfonylurea were also randomized to addition or no addition of metformin, to enable assessment of combination treatment [89,90]. The patients were followed for a mean of 11 years and the endpoints were the development of clinically significant diabetic complications. The aim was to achieve a fasting plasma glucose <6 mmol L\(^{-1}\).

Sulfonylurea treatment achieved the same glycemic control over the 10 years as did insulin treatment, with less increase in body weight [90] (median HbA1c: chlorpropamide 6.7%, glibenclamide 7.2%, insulin 7.1%, conventional 7.9%). These data are somewhat misleading since glycemic control in all groups deteriorated with time. For example in the intensively treated groups which included those randomized to sulfonylureas, median HbA1c increased for each successive 5-year treatment period (6.6% in the first 5 years, 7.5% in the second 5 years, and 8.1% in the last 5 years) [91].

A detailed analysis of the UKPDS monotherapy data showed that the percentage of patients treated with sulfonylureas that achieved a HbA1c <7% were 50% at 3 years, 34% at 6 years and 24% at 9 years [92]. An analysis of 1305 patients treated with glibenclamide or chlorpropamide over 6 years showed that loss of sulfonylurea responsiveness was related to higher initial plasma glucose concentrations at randomization, younger age of presentation of diabetes, lower initial β-cell reserve as measured by the HOMA model, and those randomized to glibenclamide compared with chlorpropamide [93]. For example, 61% of patients with a fasting plasma glucose ≥10 mmol L\(^{-1}\) required additional antihyperglycemic therapy by 6 years. By contrast only 23% required additional therapy if the fasting plasma glucose was <7.8 mmol L\(^{-1}\).

During the first 6 years of the UKPDS, β-cell function as estimated by the HOMA model was assessed in 511 patients allocated to sulfonylurea therapy and compared to that of 376 patients allocated to diet [94]. At randomization the patients on conventional therapy (diet without drugs) had 51% of normal β-cell function and the sulfonylurea patients 46%. The β-cell function of the patients on diet decreased progressively
over the 6 years to 28%. The mean β-cell function in the sulfonylurea-treated patients rose to 78% at 1 year but then decreased progressively and was 52% at 6 years. The data were interpreted as consistent with an inability of sulfonylureas to change the progressive β-cell loss that occurs in T2DM. Oral antidiabetic therapy failure most likely is the result of the progressive loss of β-cell function with time.

Sulfonylurea treatment has been shown to cause a near-euglycemic remission, which persists for at least several months and possibly years after the drugs have been discontinued. This was well documented by Singer and Hurwitz in 1967 [95] following tolbutamide (31% of patients) or chlorpropamide treatment (17% of patients), and later confirmed by Lev-Ran [96]. It is likely that the remission is related to excellent glycemic regulation rather than to a specific effect of sulfonylurea treatment. The recent studies of remission in African-American patients with T2DM [97–99] and those in Chinese patients [100] support the concept that early attainment of euglycemia by either insulin or sulfonylureas in new-onset T2DM can lead to remission of diabetes with improved insulin secretory function for several years.

A major shortcoming of sulfonylurea treatment appears to be their short durability of glycemic effect. This was initially recognized by the high secondary failure rate of sulfonylurea therapy [79], and then quantified by the UKPDS data [93,94]. A 5-year study (ADOPT) compared the durability of effect of glibenclamide to that of metformin and rosiglitazone [101]. The sulfonylurea was superior in both glucose and HbA1c lowering at 6 months, but lost most of its effectiveness by 18 months of treatment. The other drugs showed significantly greater durability of glycemic lowering beyond 1 year than glibenclamide. One mechanism by which sulfonylureas lose their effectiveness in treating hyperglycemia is postulated to be related to long-term toxic effects on the pancreatic β cells leading to an increase in apoptosis [102,103].

Patients who are most likely to show an excellent or good glycemic response to sulfonylurea therapy are those who have relatively recently diagnosed T2DM, have mild to moderate fasting hyperglycemia (fasting plasma glucose levels below 12.5 mmol L\(^{-1}\)) on dietary management, have a good β-cell reserve as reflected by high C-peptide levels, have a normal weight or are moderately obese, and have either never been treated with insulin or are taking less than 30 U of insulin daily [81,82].

Similarities and differences among sulfonylureas

All sulfonylureas appear to have the same mechanism of action on the β cell and at appropriate dose will equally stimulate insulin secretion. Differences in the clinical characteristics of individual sulfonylureas are due to variability in rates of absorption, plasma half-life, binding affinities and kinetics of interactions with the SUR binding sites, mechanisms of metabolism, and routes of excretion [5,81,82]. Side effects can be related to affinities and binding kinetics to SUR isoforms in extrapancreatic tissues or unrelated to K\(_{\text{ATP}}\) channel binding and attributable to molecular structures other than those responsible for K\(_{\text{ATP}}\) channel binding. Characteristics of the commonly prescribed sulfonylureas are detailed in the following sections.

**Tolbutamide**

Tolbutamide in various formulations from different manufacturers is in use in many countries. There are differences in absorption rate between formulations [104,105] and this may be important, as more rapid absorption may signify a higher efficacy in reducing postprandial hyperglycemia. Tolbutamide is more rapidly absorbed than chlorpropamide [6], but less rapidly than glipizide. Many, but not all, tolbutamide formulations have a high degree of absorption, but there are no appropriate estimates of the absolute bioavailability of this drug. Concomitant food intake may not delay the absorption of tolbutamide, but its efficacy may be improved if the drug is given half an hour before meals rather than with meals.

The volume of distribution of tolbutamide is small, like that of all other sulfonylureas. It is highly bound to plasma albumin, but the extent of protein binding decreases with age, roughly in parallel with decreasing plasma albumin levels [104]. Tolbutamide is completely metabolized, mainly to hydroxytolbutamide and subsequently to carboxytolbutamide, both of which have little or no activity. Metabolism occurs by oxidation, and it has been suggested that the first oxidative step to hydroxytolbutamide is genetically polymorphic with a trimodal distribution [106]. Phenylbutazone [104] and cimetidine [104] may inhibit the metabolism of tolbutamide.

Little, if any, tolbutamide is excreted unchanged. The metabolites are excreted by the kidney [104].

The usual dosage of tolbutamide is 0.5–1.0 g three times daily.

**Chlorpropamide**

Chlorpropamide is more slowly absorbed than tolbutamide and has the longest elimination half-life of all sulfonylureas. This leads to pronounced accumulation with very small fluctuations over the day within the individual [107]. Therefore, the elimination rate is the major determinant of its effect, whereas variations in its rate of absorption are unimportant. Hence the timing of the daily dose is irrelevant, and it is also irrelevant whether chlorpropamide is ingested before or with breakfast. There are no data on the absolute bioavailability of chlorpropamide. Like that of all other sulfonylureas, the volume of distribution is small, but its binding to plasma albumin is less pronounced. Chlorpropamide undergoes extensive, albeit slow, metabolic transformation. This occurs by oxidation to 2-hydroxylchlorpropamide, 3-hydroxychlorpropamide, and p-chlorobenzene sulfonylurea [104]. It is not clear whether the hydroxylated metabolites are active; however, they are more
rapidly eliminated than the parent drug, at least in subjects with normal renal function.

A significant portion of chlorpropamide is excreted unchanged by the kidneys [104]; this proportion can be increased by alkalization and reduced by acidification of the urine [6,104]. The elimination half-life of chlorpropamide is long but variable (24–48 h), partly because of variations in the excretion rate of the unchanged compound and partly because of variations in the metabolic turnover rate. This also helps to explain why there is a very large inter-individual variation in the steady-state plasma concentrations of chlorpropamide during treatment [107].

The usual dosage of chlorpropamide is 250–500 mg once daily.

Chlorpropamide is more commonly associated with long-lasting hypoglycemic events than are most other sulfonylureas, possibly with the exception of glibenclamide (see later); this is probably related to the slow elimination and pronounced accumulation of the drug. On the other hand, chlorpropamide is the only sulfonylurea with elimination that can be enhanced by forced diuresis and urine alkalization [6,104]. The pronounced accumulation may also help to explain why chlorpropamide, more often than other sulfonylureas, is involved in events of ethanol-provoked facial flushing.

A major unique side effect of chlorpropamide treatment is water retention and hyponatremia. This is related to stimulating the release of vasopressin from the posterior pituitary and to potentiating vasopressin action on the renal tubule [76]. This side effect might explain the observation in the UKPDS that the patients randomized to chlorpropamide had significantly higher systolic and diastolic blood pressures throughout the study than those assigned to glibenclamide, insulin, or conventional treatment [91]. Alcohol-induced facial flushing is a complication of chlorpropamide treatment in some patients with T2DM [108]. The mechanism is related to an inhibition of aldehyde dehydrogenase and elevated blood acetaldehyde levels [108]. Alcohol-induced flushing is not unique to chlorpropamide and occurs with a lower frequency with tolbutamide and only rarely with glibenclamide.

The UKPDS clearly showed that chronic chlorpropamide treatment achieves better glycemic control than does glibenclamide (median 10-year HbA1c: 6.7% vs. 7.2%; p < 0.008) [91]. Despite the better glycemic control, the chlorpropamide-treated patients had less risk reduction for the progression of retinopathy than did the glibenclamide-treated group [91].

**Glibenclamide**

The most extensively used sulfonylurea in most parts of the world is glibenclamide (called glyburide in the United States). It was the first “second generation” sulfonylurea to be introduced in clinical practice, and it is one of the most potent. Both nonmicronized and micronized generic formulations have appeared, but it is not clear whether they are fully equivalent to the original products. Although concomitant food intake does not seem to delay the absorption of glibenclamide, its efficacy may be increased if given before meals, at least in the short term. Indeed, 2.5 mg given half an hour before breakfast was more effective than 7.5 mg given together with breakfast [109]. The rate of absorption varies extensively between individuals [110].

Like other sulfonylureas, glibenclamide has a small volume of distribution, and it is very extensively bound to albumin.

Glibenclamide is completely metabolized, mainly by oxidation to two hydroxylated metabolites, both of which have blood glucose reducing activity [111].

About half of a given dose of glibenclamide is eliminated by renal excretion of metabolites, and a significant portion is excreted through the bile. Although renal insufficiency should not alter the elimination rate of this very lipophilic and completely metabolized drug, it may nevertheless increase the risk of hypoglycemia [104]. This may be because renal insufficiency is associated with reduced drug binding to albumin; in addition, renal impairment may lead to accumulation of polar, active metabolites [104].

Both clinical experience and various studies support the view that glibenclamide should be classified as a long-acting sulfonylurea. Indeed, glibenclamide has been associated with long-lasting (and hence dangerous) hypoglycemic reactions more often than other sulfonylureas except, perhaps, chlorpropamide [3,104,112–114].

The usual daily dose of glibenclamide is 1.75–14 mg (micronized) or 2.5–20 mg (nonmicronized), given in a single morning dose or split in a morning and an evening dose.

Glibenclamide and glipizide have been compared. Glipizide releases insulin more rapidly than glibenclamide [112–116]; however, glibenclamide suppresses hepatic glucose output more than does glipizide [115].

In a placebo-controlled study wherein glibenclamide and glipizide were given in a crossover fashion for 6 months, and in which most patients were continuously exposed to either drug, there was no difference in the final mean daily dose, or in final mean blood glucose or hemoglobin levels. Although fasting blood glucose values tended to be lower in patients on glibenclamide and postprandial blood glucose values lower in those on glipizide, neither difference was significant [104].

Despite concerns about the use of sulfonylurea drugs in pregnancy, a comparison of glibenclamide and insulin treatment of 404 women with gestational diabetes showed that both treatments gave equal glycemic control and no difference in maternal and neonatal complications. Thus glibenclamide may be a clinically effective alternative to insulin treatment in women with gestational diabetes [117,118].

**Gliclazide**

Gliclazide is a rapid-acting and short-acting sulfonylurea that is widely used in some countries, including France and the United Kingdom. Because of the aminoazabicyclo-octyl ring in its chemical structure, gliclazide is thought to have additional benefits in that it can scavenge reactive oxygen species [119].
It is rapidly absorbed, but its absolute bioavailability is not known. Absorption is slower in the elderly [120,121]. Like other sulfonylureas, gliclazide is extensively bound in plasma and has a small volume of distribution [104,121]. It is not known whether intake before meals makes gliclazide more effective than does intake together with meals.

Gliclazide is almost completely metabolized, and the metabolites seem to be inactive [104].

Most gliclazide is eliminated by excretion of the inactive metabolites through the kidneys, but a small fraction is eliminated by the bile [104]. The elimination half-life is short but variable, between 6 and 15 h. Although this might not be expected for a completely metabolized drug, the elimination half-life may be increased in subjects with renal insufficiency [104].

It has been argued that gliclazide offers distinct advantages over other sulfonylureas. Its antioxidant activity is reported to protect pancreatic β cells from the destructive effects of oxidative stress [122,123]. It has been claimed that gliclazide is able to reduce platelet aggregation [124]. However, it is most likely that all sulfonylureas are able to improve platelet aggregation as an indirect effect of improved blood glucose control [6]. There are no comparative studies showing that gliclazide is superior to other sulfonylureas in reducing platelet aggregation. In terms of blood glucose control, gliclazide seems to be as effective as other sulfonylureas [104].

The usual daily dose is 40–320 mg, given in a single morning dose or in the morning and evening.

Gliclazide MR is a modified-release formulation using a hydrophilic matrix allowing a once-daily dosing regimen with optimal and consistent release of gliclazide [125]. The usual dose is 30 to 120 mg once daily. The GUIDE study which compared gliclazide MR to glimepiride showed equivalence in improving glycemic control but 50% fewer confirmed hypoglycemic events with gliclazide MR [126]. Gliclazide MR was the primary agent used in the intensively controlled cohort in the ADVANCE study which was a study to determine the effect of intensive glycemic control compared to ordinary glycemic control on combined microvascular and macrovascular complications in 11,400 type 2 diabetic patients treated for a median of 5 years [127]. The results showed a decrease in early microvascular endpoints. Avogaro has reviewed those data and additional studies that showed the gliclazide has general antioxidant and anti-inflammatory effects [128]. Despite the emphasis on the antioxidant and anti-inflammatory actions of gliclazide, there are no convincing studies that these effects translate into better clinical outcomes in patients with T2DM than other sulfonylureas.

Glimepiride

Glimepiride is widely used in the United States and in several European countries. It is the most rapidly acting sulfonylurea available, and is also one of the most short-acting [6,104]. Its potency and intrinsic activity are in the same range as those of glibenclamide. The absorption and bioavailability of glimepiride are fast and complete [6,129]. It is more rapidly absorbed when taken before breakfast than when ingested together with breakfast, and intake before breakfast is also associated with a more appropriate timing of insulin release relative to the meal, and with an enhanced efficacy of the drug [112]. Its absorption rate correlates with its efficacy [130], and pronounced hyperglycemia may reduce the absorption rate [131].

Glimepiride has been shown to improve the acute insulin release in response to a meal [130], and this capacity may be maintained during long-term therapy, at least when the exposure is continuous [132,133]. This is interesting in view of the possibility that chronic, continuous sulfonylurea exposure may desensitize the β cell to sulfonylurea stimulation [6,47,133].

The volume of distribution of glimepiride is small, like that of other sulfonylureas, and the binding to albumin is very extensive [104].

Glimepiride is completely metabolized, mainly by oxidative hydroxylation [104]. The metabolites appear to be inactive [104].

Glimepiride is rapidly eliminated with a half-life of 1–5 h. The half-life does not seem to increase with age. Renal insufficiency does not seem to alter the elimination rate of glimepiride. These facts may explain why glimepiride seems to carry a lower risk of long-lasting hypoglycemia than do glibenclamide and chlorpropamide.

Although glimepiride is rapidly eliminated, a morning dose of 7.5 mg or more maintains effective plasma concentrations for more than 12 h [6,133,134]. This may help to explain why glimepiride, despite its short half-life, can be equally effective when given once daily or three times daily [133,134]. On the basis of available documentation, dosage should be in the range of 2.5–10 mg daily, either as a single morning dose half an hour before breakfast or divided in a morning and an evening dose. In the United States, the official maximum dose is 40 mg, but there is very limited support for increased efficacy with daily doses over 10–15 mg. Instead, dose increase from 15 to 25 mg d⁻¹ has been found to impair rather than improve glucose control [73], and a placebo-controlled study using 3-month periods of glimepiride at 10, 20, and 40 mg daily showed impaired glucose and insulin responses above 10 mg daily [74]. On the other hand, ethnic differences may exist in the disposition and effect of sulfonylureas, and hence in appropriate dosage.

Glimepiride

Glimepiride is a second generation sulfonylurea that differs structurally from glibenclamide in its two side chains. It is rapidly and completely absorbed after oral administration. Peak serum concentrations are achieved at 2–3 h [135]. When glimepiride is taken with meals the time to reach the maximum concentration and the concentration achieved are decreased by 10% [135]. Plasma protein binding of glimepiride is >99% and the volume of distribution is small. The elimination half-life is 5–9 h [135]. Glimepiride is completely biotransformed by oxidative metabolism (CYP2C9) in the liver and
Repaglinide is a short-acting insulin secretagogue and is usually administered in a dose of 0.5–4.0 mg before each meal [142,143]. Because of its pharmacologic properties, its primary action is to lower postprandial hyperglycemia [142,143]. Despite its short duration of action, repaglinide has a significant effect in lowering fasting hyperglycemia [142,143]. Repaglinide is useful as monotherapy or in combination with metformin or other antidiabetic agents. Its short duration of action and administration prior to each meal makes it an ideal treatment for type 2 diabetic patients with a flexible lifestyle [142,143].

Repaglinide monotherapy has been shown to give the same improvement in glycemic control as glimepiride or glyburide in 12-month studies in type 2 diabetic patients [142,143]. Rates of hypoglycemia with repaglinide were comparable to those seen in comparator trials versus glipizide and less than those versus glyburide [142,143]. Weight gain on repaglinide therapy (2.45 kg yr^{-1}) is less than that occurring with glibenclamide (3.64 kg yr^{-1}) [143]. Repaglinide treatment has not been associated with an increase in cardiovascular events or mortality despite the nonselectivity of its binding to SUR isoforms [142]. Insufficient data are available to compare durability of effect or β-cell survival of repaglinide to the sulfonylureas.

Adverse effects of sulfonylureas

In the past, the frequency of side effects of sulfonylureas were thought to be low, in the range of 3–5% and most adverse effects to be mild and reversible upon withdrawal of the drugs. More recent data from retrospective data analyses and large healthcare databases have revised our opinions and particularly have shown significant complication rates in subsets of the T2DM population. The most serious side effect of insulin secretagogues is hypoglycemia. Weight gain is a modest problem, but its clinical impact can only be estimated. An ongoing controversial issue which persists even after 40 years is whether some sulfonylureas may cause unfavorable cardiovascular effects during episodes of myocardial ischemia.

Hypoglycemia

The incidence of hypoglycemia in patients with T2DM has been estimated to be between 12 and 30%. While mild hypoglycemic events are uncomfortable and worrisome for the patient, it is the major or severe hypoglycemic events (symptoms requiring assistance by a third party) which cause the increased morbidity and mortality and the extensive healthcare costs. The magnitude of the problem of sulfonylurea-induced severe hypoglycemia varies remarkably among the different published reports. Some investigators claim that it is rare and not a major clinical problem, while most others find it to be a frequent and serious complication. In reviewing the available data, it appears that reports from single centers with a limited number of type 2 diabetic patients are less likely to view the risk as serious, while studies from large clinical trials and healthcare databases find severe hypoglycemia to be a significant complication of sulfonylurea therapy.
In the UKPDS study, the rates of any hypoglycemic symptoms were 11.0% for chlorpropamide, 17.7% for glibenclamide, 36.5% for insulin, and 1.2% for lifestyle management. The rates of major (severe) hypoglycemia were 0.4% for chlorpropamide, 0.6% for glibenclamide, 2.3% for insulin, and 0.1% for lifestyle [91]. A US population-based retrospective analysis of the incidence of severe hypoglycemia in older individuals reported rates of 1.23 per 100 person-years with sulfonylurea use versus 2.76 per 100 person-years in insulin users [144]. A retrospective study of 33,243 clinical practice patients taking sulfonylures in the United Kingdom reported an annual hypoglycemia rate of 1.8% [145]. The rate for those less than 65 years was 1.4% per year and for those 65 years or older 2.0% per year. The rate for glibenclamide was greater than for other sulfonylures.

Within a large integrated healthcare system (Kaiser Permanente), 985 (10.8%) of 9094 eligible type 2 diabetic survey respondents, reported having had one or more severe hypoglycemic episodes requiring assistance in the preceding year [146]. Of the entire cohort, 22.4% were being treated with insulin, 54.8% were taking an insulin secretagoue, and 22.4% were on other antidiabetic treatments. Severe hypoglycemia was reported in 19% of the insulin-treated patients and 9.6% of the patients taking an oral secretagoue. Severe hypoglycemia occurred as frequently in those with poor glycemic control (HbA1c $> 9\%$) as it did in those achieving near normal glycemic control ($7.0 – 7.9\%$). In the DARTS/MEMO study from Tayside, Scotland data collected over a 12-month period on severe hypoglycemia that required emergency medical treatment were determined in 8655 patients with diabetes [147]. A total of 244 episodes were recorded in 160 patients comprising 69 (7.1%) with T1DM, 66 (7.3%) with T2DM treated with insulin, and 23 (0.8%) with T2DM treated with sulfonylures. One in 3 was treated solely by the ambulance service, while the remainder were treated in the hospital. The total estimated cost of emergency treatment for severe hypoglycemia was $\leq 692,078$ in 1 year. An observational, intensively monitored 9–12-month study in the United Kingdom assessed hypoglycemia in three treatment groups of patients with T2DM (sulfonylurea, insulin for $< 2$ years, and insulin for $> 5$ years) and two treatment groups of patients with T1DM ( $< 5$ years and $> 15$ years duration of disease) [148]. The incidence of severe hypoglycemia was the same in patients with T2DM treated with sulfonylures ($7\%$) as in those treated with insulin $< 2$ years ($7\%$). The incidence calculated to 0.1 and 0.2 episodes per patient-year. It has been estimated that each year $> 5000$ patients in the United Kingdom will experience a severe hypoglycemic reaction caused by their sulfonylurea therapy which will require emergency intervention [149]. Each hospital admission for severe hypoglycemia in 2008 was estimated to cost £1000.

The prevalence of sulfonylurea-associated hypoglycemia requiring hospital treatment has been reported as 0.38 per 1000 treatment-years in Switzerland, 0.19 per 1000 treatment-years in Sweden [150], and 4.2 per 1000 treatment-years in the Swedish island of Gotland, where the use of sulfonylurea is particularly common [114]. The Swiss survey reported a fatality rate of 4.3% among the hospital-admitted patients [113]. In a US survey done in the mid 1970s, 10% of hypoglycemic patients referred to hospitals died, and 3% had permanent neurologic damage [151].

Long-lasting, and hence serious, hypoglycemia occurs more often with long-acting sulfonylures, such as glibenclamide and chlorpropamide, than with short-acting ones, such as glipizide and tolbutamide [113,114,145,152]. The numbers of serious cases in the Swiss review were 0.38, 0.34, 0.15, and 0.07 per 1000 treatment-years for glibenclamide, chlorpropamide, glipizide, and tolbutamide, respectively [113]. Figures from Sweden for the period 1975–1985 show that the number of long-lasting hypoglycemic cases per million defined daily doses were 0.195 for glibenclamide and 0.184 for chlorpropamide, but only 0.004 for glipizide and 0.072 for tolbutamide [150]. Data from the retrospective elderly type 2 diabetic population in the United States found an incidence of hypoglycemia requiring medical intervention to be greatest with glibenclamide (1.66 episodes/100 person-years) and least with tolbutamide (0.35 episodes/100 person-years) [153]. Chlorpropamide use was associated with an event rate similar to that for glibenclamide, and glipizide use had an event rate that was approximately half that of glibenclamide. A recently reported Swiss study found the incidence of severe hypoglycemia to be threefold greater for long-acting sulfonylures as compared to short-acting forms [152]. As noted earlier the incidence of severe hypoglycemia in type 2 diabetic patients treated with glimepiride is significantly less than those treated with glibenclamide (0.86 episodes/1000 person-years vs. 5.6/1000 person-years) [154].

A predisposing factor for the development of sulfonylurea-induced hypoglycemia is age: most serious cases in Sweden were patients over 75 years old, 21% of whom were over 85 years old [150]. Other predisposing factors are reduced food intake, intercurrent illness, renal disease, hepatic disease, and cardiovascular disease [113,150,155]. It is of note that patients with mutations in the hepatic nuclear factor 1 alpha gene which is the most common form of maturity-onset diabetes of the young (MODY) are exquisitely sensitive to the hypoglycemic effects of sulfonylures. Treatment of these patients with sulfonylures is extremely effective; however, the risk of developing severe hypoglycemia is high [156]. Therefore treatment of these patients should be initiated with low doses of short-acting sulfonylures.

There is no apparent dose relation to the development of severe sulfonylurea-induced hypoglycemia in type 2 diabetic patients: severe and fatal hypoglycemia may occur with low doses of the offending sulfonylurea. Drug interactions frequently may be involved, often involving aspirin, salicylate, or alcohol (see later).

The consequences of severe hypoglycemia include both increased morbidity and mortality and a significant increase in healthcare costs. Patients with diabetes (78.7% type 2) who reported one or more severe hypoglycemic reactions had a
3.4-fold increase in subsequent 5-year mortality compared to those diabetic patients who reported no severe hypoglycemic episodes [157]. These findings are consistent with the increase in cardiovascular morbidity and mortality reported in type 2 diabetic patients in the ADVANCE trial who experienced one or more severe hypoglycemic episodes [158].

Current US data show that the insulin and oral hypoglycemic agents were the second and fourth most common cause of drug-related emergency hospitalizations in patients aged ≥ 65 years for the period 2007–2009 [159]. The estimated annual hospitalizations for sulfonylurea and insulin-induced severe hypoglycemia were 10,656 and 13,854, respectively (10.7% and 13.9% of all drug-related admissions). Of the patients seen in emergency departments for sulfonylurea-induced hypoglycemia 51.8% required hospital admission. In Korea, 132 of 320 (41.3%) patients treated in an emergency department from 2006–2009 for severe hypoglycemia were taking sulfonylureas as contrasted to 29% being treated with insulin [160]. A similar study in Israel of 102 drug-induced hypoglycemic comas in patients with T2DM found that approximately 60% were due to sulfonylurea treatment [161]. Death occurred in 5 of the 102 patients. Risk factors for sulfonylurea-induced hypoglycemic coma were age older than 60 years, renal dysfunction, decreased intake of energy, infection, and concomitant use of drugs that potentiate hypoglycemia [161].

The recent emphasis on intensive glycemic control has caused a decrease in hospitalizations for hyperglycemia but has increased the hospitalizations for hypoglycemia. An analysis of data from Medicare beneficiaries 1999–2011 calculated the change in hospital admissions for hyperglycemia and hypoglycemia in diabetic patients. From 1999 to 2011 the rate of hospital admission for hyperglycemia fell from 820 to 676 per 100,000 patient-years. During the same time frame, the rate of hospitalization for hypoglycemia went from 367 to 612 per 100,000 patient-years [162].

A vicious cycle of dementia and hypoglycemia is being recognized. A higher risk of hypoglycemia occurs among diabetic patients with dementia (adjusted hazard ratio 3.1) and a higher risk of dementia occurs among diabetic patients with recurrent episodes of hypoglycemia (adjusted hazard ratio 2.1) [163]. The associated with intensification of glucose control with insulin and/or sulfonylurea therapy has focused attention on the fact that goals and types of therapy for many patients with diabetes have been excessive [164].

The impact of the increasing rate of hypoglycemia in patients with T2DM has multiple effects. Severe hypoglycemia adds an additional cost dimension to diabetes care. A 2002 analysis from Germany estimated that the cost of severe hypoglycemia amounted to $44,338/100,000 inhabitants in type 2 diabetic patients versus $8129/100,000 inhabitants in type 1 diabetic patients [165]. A study of the incidence and costs of severe hypoglycemia requiring attendance by emergency medical services in South Central England estimated the annual cost of emergency calls for severe hypoglycemia to be £13.6 million for England [166]. Other consequences of severe hypoglycemia are that fear of hypoglycemia and hypoglycemia itself is a significant barrier to good glycemic control.

The occurrence of severe sulfonylurea-induced hypoglycemia is minimized by avoiding sulfonylureas, particularly long-acting ones, in patients with predisposing conditions or who are taking potentially interacting drugs. When such cases do occur, the patients should be hospitalized. A bolus of 50% glucose should be given intravenously and should be followed by continuous infusion of 10% or 20% glucose. Blood glucose levels should be monitored for at least 3 days and maintained at 6–8 mmol L⁻¹ (110–145 mg dL⁻¹). If intravenous glucose treatment is insufficient, hydrocortisone and/or glucagon administration may be useful. A somatostatin analogue (octreotide) has been shown to reverse sulfonylurea-induced hyperinsulinemia and hypoglycemia and has been shown to be very effective in severe or intractable cases of sulfonylurea-induced hypoglycemia [167–169].

**Weight gain**

Weight gain is a frequent complication of sulfonylurea treatment. The magnitude of weight gain differs among the different sulfonylureas and the particular clinical studies. Older, well-controlled studies found that the mean increase in body weight in patients with T2DM on sulfonylurea treatment was 2.8 kg [170]. This contrasted with the mean weight change ranging from −2.9 to +1.5 kg that occurred with comparable improvement in glycemic control by metformin treatment [170]. The UKPDS provides the best long-term data on body weight changes with sulfonylurea treatment. Glibenclamide treatment resulted in a mean 4 kg increase in body weight in the first 4 years of treatment [94] which then plateaued for the remaining 7 years of the study [91]. The diet-treated control patients had a mean 1 kg weight increase during the 11 years of the study. In contrast, metformin-treated patients had the same mean body weight increase as the diet-treated control patients and the insulin-treated patients had a progressive increase in body weight throughout the study achieving a body weight gain of 8 kg [91]. The ADOPT study carried out 20 years later confirmed that glibenclamide weight gain occurs in the first 1–2 years of treatment (1.6 kg) and then plateaus [101].

Epidemiologic data have shown that all-cause mortality is related to body mass index (BMI). The hazard ratios for all-cause mortality in overweight and obese diabetic patients relative to patients with BMI 18.5–25 kg m⁻² are: overweight 0.97, type 1 obese (BMI 30–34.9 kg m⁻²) 1.13, and type 2 and 3 obese (BMI ≥ 35 kg m⁻²) 1.48 [171]. Voluntary weight loss has been demonstrated to decrease all-cause mortality. There are no specific data to prove that sulfonylurea-induced weight gain has detrimental effects on outcomes in patients with T2DM. There are several potential mechanisms that have been proposed to explain the weight gain associated with sulfonylurea treatment. Improvement in glycemic control decreases glycosuria and can increase caloric balance if food intake is
not sufficiently decreased. Many patients taking sulfonylureas have episodes of mild hypoglycemia that manifest themselves as hunger and contribute to excess calorie intake. There is some evidence that hyperinsulinemia itself may increase appetite.

**Cardiovascular safety**

The most frequently debated adverse effect of sulfonylureas is that of an alleged increase in cardiovascular mortality. This was initiated in 1970 by the University Group Diabetes Program (UGDP) report that subjects treated with tolbutamide had an increase in cardiovascular mortality compared to those treated with placebo [172]. After extensive analyses and debates carried out over several years, it seemed that these data might have been based on a spuriously low mortality rate in the placebo group [173]. However, subsequently several potential mechanisms have been elucidated which could account for sulfonylureas having detrimental effects during cardiac ischemia and the cardiovascular safety of sulfonylureas has continued to be an unresolved issue.

The discovery of $K_{\text{ATP}}$ channels in the myocardium and vascular smooth muscle led to an understanding of their importance in the regulation of cardiac physiology. Myocardial ATP levels are ordinarily high because of the high rate of substrate oxidation. The $K_{\text{ATP}}$ channels are therefore closed resulting in increased calcium ion entry into the myocardium and increased sarcolemma contraction. During myocardial ischemia, ATP levels fall causing the $K_{\text{ATP}}$ channels to open. This results in decreased calcium entry into the myocardium, decreased energy expenditure, a reduction in action potential duration and an increase in re-entrant arrhythmias. Closure of the coronary artery smooth muscle cell $K_{\text{ATP}}$ channel causes compensatory vasodilatation.

A transient episode of myocardial ischemia produces metabolic changes which protect the myocardium from the effects of a subsequent more prolonged episode of ischemia. This phenomenon is called ischemic pre-conditioning and reduces the myocardial damage caused by the prolonged ischemia [174,175]. It appears that ischemic preconditioning is related to opening of the cardiac mitochondrial $K_{\text{ATP}}$ channel.

The SUR subunit of the myocardial $K_{\text{ATP}}$ channel is the SUR2A isoform, that of the coronary artery is the SUR2B isoform, and that of the mitochondria is SUR1 but coupled to Kir6.1. Sulfonylurea drugs in the doses used clinically could close the myocardial, coronary artery smooth muscle cell and mitochondrial $K_{\text{ATP}}$ channels and prevent the protective adaptation of the heart to anoxia and ischemia if they had sufficient affinities for the SUR2A, SUR2B, and Kir6.1 isoforms. The only sulfonylurea that has been shown to be capable of this in humans is glibenclamide. Ordinary pharmacologic doses of glibenclamide block ischemic preconditioning in both animal and human models of myocardial ischemia [139,140]. Other sulfonylureas such as glimepiride do not [139].

A second factor which could cause cardiac safety issues in diabetic patients taking sulfonylureas is effects of severe hypoglycemia on electrical impulse transmission in the myocardium leading to arrhythmias. Several studies have shown that severe hypoglycemia leads to significant prolongation of the QT interval [176] which is known to predispose to arrhythmias.

Whether these potential effects of sulfonylureas lead to increased myocardial damage and/or increased cardiovascular events can only be determined by clinical data or preferably controlled, clinical trials. After 40 years the data on sulfonylureas and cardiac safety are still controversial [177–181]. Retrospective analyses of several large databases report that chronic sulfonylurea treatment of patients with T2DM is associated with higher all-cause or cardiovascular mortality than patients in the database who were treated with metformin (Table 42.2). The issue of course is whether sulfonylureas increase cardiovascular mortality or metformin decreases cardiovascular mortality. Some of the studies summarized in Table 42.2 that compared the cardiovascular safety of the various sulfonylureas to each other show that glibenclamide is more detrimental to the heart than the other sulfonylureas. A study of the outcomes of 1310 diabetic patients included in the French Registry of acute myocardial infarctions in 2005 found that patients being treated with glibenclamide had an in-hospital mortality of 7.5% compared to one of 2.7% of those patients being treated with glipizide or glimepiride [192]. In contrast, a meta-analysis of 21 studies from 1995–2005 reported that glibenclamide caused more hypoglycemia (83%) but the same rates of risk of cardiovascular events or death compared to other insulin secretagogues [193]. A retrospective cohort analysis using an academic health center enterprise-wide electronic health record of 11,141 patients with T2DM compared the risk of overall mortality of patients initiated on monotherapy with glyburide (4279), glipizide (4325), or glimepiride (2537) found a trend toward greater mortality of those in the glipizide (hazard ratio 1.39) and glyburide (hazard ratio 1.36) treated groups as compared to the glimepiride group [194].

In summarizing the large amount of data available, it does not appear that there is sufficient evidence to conclude that sulfonylureas in general cause an increase in cardiovascular events. Many of the studies were in diabetic populations with acute myocardial infarction or acute coronary syndrome so that detrimental effects of sulfonylureas on the cardiovascular system should have been evident. There continues to be concern, however, that glibenclamide in some subsets of patients may cause an increase in mortality. Since glibenclamide has never been shown to be more beneficial in controlling hyperglycemia than other sulfonylureas, and has a higher rate of severe hypoglycemia as well as possible detrimental cardiovascular effects, there is no rationale for its use as a therapeutic agent in T2DM. Metformin therapy seems to decrease cardiovascular events and mortality.
Table 42.2 Analyses of sulfonylurea effects on cardiovascular events and/or mortality in patients with type 2 diabetes

<table>
<thead>
<tr>
<th>Sulfonylurea</th>
<th>Comparator</th>
<th>Number</th>
<th>Duration of follow-up</th>
<th>Results ratio (95% CI)</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sulfonylureas</td>
<td>Metformin</td>
<td>5730</td>
<td>Up to 8 yrs</td>
<td>RR CV mortality 3.71 (2.64–5.22)</td>
<td>[182]</td>
</tr>
<tr>
<td>Sulfonylureas</td>
<td>Metformin</td>
<td>422 CHF</td>
<td>1 yr</td>
<td>RR all-cause mortality 1.69 (1.04–2.78)</td>
<td>[183]</td>
</tr>
<tr>
<td>Sulfonylureas</td>
<td>Metformin</td>
<td>98,665/155,025</td>
<td>Composite CVD events</td>
<td>1.21 (1.13–1.30)</td>
<td>[184]</td>
</tr>
<tr>
<td>Glyburide</td>
<td>Metformin</td>
<td>3469/711 MI</td>
<td>2.2 yrs</td>
<td>HR composite CVD events 1.31 (1.17–1.46)</td>
<td>[185]</td>
</tr>
<tr>
<td>Glyburide</td>
<td></td>
<td></td>
<td></td>
<td>1.19 (1.06–1.32)</td>
<td></td>
</tr>
<tr>
<td>Glyburide</td>
<td></td>
<td></td>
<td></td>
<td>1.25 (1.11–1.42)</td>
<td></td>
</tr>
<tr>
<td>Glyburide</td>
<td></td>
<td></td>
<td></td>
<td>1.18 (1.03–1.34)</td>
<td></td>
</tr>
<tr>
<td>Glipizide</td>
<td></td>
<td></td>
<td></td>
<td>1.03 (0.88–1.22)</td>
<td></td>
</tr>
<tr>
<td>First generation sulfonylurea</td>
<td>Metformin</td>
<td>304 CAD Prospective study</td>
<td>5 yrs</td>
<td>HR composite CVD events 1.85 (1.11–3.33)</td>
<td>[186]</td>
</tr>
<tr>
<td>Glyburide</td>
<td>High-dose</td>
<td></td>
<td></td>
<td>2.1 (1.0–4.7)</td>
<td>[187]</td>
</tr>
<tr>
<td>Metformin</td>
<td></td>
<td></td>
<td></td>
<td>1.3 (1.2–1.4)</td>
<td></td>
</tr>
<tr>
<td>Glyburide</td>
<td></td>
<td></td>
<td></td>
<td>0.8 (0.7–1.1)</td>
<td></td>
</tr>
<tr>
<td>Glyburide</td>
<td>Metformin</td>
<td>271/129 Emergent PCI</td>
<td>Up to 9 yrs</td>
<td>aHR CV mortality 2.91 (1.26–6.72)</td>
<td>[189]</td>
</tr>
<tr>
<td>Sulfonylureas</td>
<td>First-line treatment</td>
<td>Metformin</td>
<td>15,687/76,811</td>
<td>2.9 yrs</td>
<td>HR all-cause mortality 1.58 (1.48–1.68)</td>
</tr>
<tr>
<td>Second-line treatment</td>
<td>DPP-4 inhibitors</td>
<td>33,983/7854</td>
<td>2.9 yrs</td>
<td>HR all-cause mortality 1.35 (1.08–1.71)</td>
<td></td>
</tr>
<tr>
<td>Glyburide</td>
<td>Gliclazide</td>
<td>1690/985 Acute MI or coronary intervention</td>
<td>2 yrs</td>
<td>Composite CV death, MI and CHF RR 1.01 (0.86–1.18)</td>
<td>[191]</td>
</tr>
</tbody>
</table>
| Metformin | Sulfonylureas | 39,721 | | Adjusted odds ratio for all-cause mortality 0.87 (0.68–1.10) | **

**Presented by Currie C at EASD meeting in Barcelona, Spain, September 2013.

**Sulfonylureas as current therapeutic agents for type 2 diabetes**

With the availability of many classes of agents to treat the hyperglycemia of patients with T2DM, it is important to examine the positive and negative aspects of sulfonylurea treatment for these patients. Sulfonylureas are glucose-independent insulin secretagogues, which means that their use will always be associated with a significant incidence of hypoglycemia with its potential moderate and severe clinical complications. Their use will lead to some weight gain, the clinical significance of which is unknown. The risks of an increase in cardiovascular events and cancer remain unclear. Long term, the lack of durability of glycemic effect and the possibility of detrimental effects on the β cell are major concerns.

Sulfonylureas have been shown to improve glycemic control when used as primary therapy in patients with T2DM inadequately controlled by lifestyle modification and when added as second- or third-line therapy to patients inadequately controlled with metformin, thiazolidinediones, insulin, and DPP-4 inhibitors. They have been used in combination with GLP-1 receptor agonists and SGLT-2 inhibitors. However, just as in first-line pharmacologic therapy, their use when combined with other agents is always accompanied by increased rates of moderate and severe hypoglycemia.

The question to be raised then is, are other strategies that are as effective and yet safer with less hypoglycemia and a potentially lower risk of CVD available to replace sulfonylureas? In this context, we turn to the insulin secretagogues which are glucose-dependent and cause little or no hypoglycemia. Obviously, the incretin-based therapies can provide such an alternative. Randomized, controlled clinical trials comparing sitagliptin to glipizide [195], saxagliptin to glipizide [196], and vildagliptin to glimepiride [197] in patients with...
Table 42.3 Genetic variants of the pancreatic β-cell that are or may be specific targets for sulfonylurea treatment

| 1 | Mutations of ABCC8 or KCNJ11 genes of the KATP channel that block the ability of ATP/ADP to regulate KATP channel closure and stimulate insulin secretion (neonatal diabetes mellitus) [204]. Ninety percent are able to be controlled on high doses of sulfonylureas rather than insulin [205]. |
| 2 | HNF-1α mutations causing MODY-3 [202,203]. Five times more sensitive to sulfonylureas than metformin. |
| 3 | Mutations of KATP channel genes that are associated with some forms of T2DM [208]:
| (a) | KCNJ11 Kir6.2 K23 variant is present in approximately 20% of Caucasians with T2DM |
| (b) | K23 variant is associated with T2DM in almost all ethnic populations tested |
| (c) | K23 variant is closely coupled with SUR ABCC8 variant A1369 |
| (d) | Haplotype A1369 ABCC8 and K23 KCNJ11 variant is responsible for increased risk for T2DM. |
| 4 | Polymorphisms of transcription factor 7-like 2 L (TCF-7L2) [206,207]. |

T2DM inadequately controlled on metformin treatment, have demonstrated that DPP-4 inhibitors are equally effective as sulfonylureas in decreasing HbA1c, yet cause no weight gain and are accompanied by rates of moderate hypoglycemia that are 10- to 15-fold lower than sulfonylureas and with almost no episodes of severe hypoglycemia. GLP-1 receptor agonists such as exenatide twice daily or once weekly as well as liraglutide are more potent in lowering glycemia than sulfonylureas or insulin, cause little or no hypoglycemia and are associated with weight loss [198–200]. While there are some concerns about incretin-based therapies being associated with an increased risk of acute pancreatitis and pancreatic cancer the current available data do not support such a concern [201]. The one clear advantage of sulfonylureas is their very low cost. One can then make the argument that DPP-4 inhibitors should be used preferentially to sulfonylureas in patients who are at increased risk of hypoglycemia from sulfonylureas. In patients in whom cost is critical, sulfonylureas are better than no therapy.

It is important, however, to recognize that sulfonylureas or meglitinides are the only agents that can close the pancreatic β-cell KATP channel when it is insensitive to closure by ATP/ADP. Such abnormalities are being recognized in individuals with abnormalities of genes for pancreatic β-cell ATP generation [202,203] or the subunits of the KATP channel [204,205]. Table 42.3 lists some of the known and potential abnormalities that can be uniquely benefitted by sulfonylurea therapy [202–208]. The glucose-dependent insulin secretion of pancreatic β-cells act through cyclic AMP and protein kinase A and do not act on the KATP channel.

Interactions

Several drugs interfere with the efficacy of sulfonylureas, by influencing their pharmacokinetics or pharmacodynamics, or both. As the majority of type 2 diabetes patients are elderly subjects or are in late middle age, they are liable to be exposed to other medication in addition to sulfonylurea. Thus, the risk of interactions is considerable. Two of the most important interactions may occur with alcohol and with aspirin, both of which may provoke, prolong, and/or deepen a hypoglycemic reaction [6,104,152]. Plausible and/or documented interactions are listed in Table 42.4. It should be noted that the degree and

Table 42.4 Mechanisms of drug interactions with sulfonylureas

| Agents that augment sulfonylurea action and may cause profound hypoglycemia: |
| (A) | Displacers of sulfonylureas from albumin binding sites |
| (1) | Aspirin and salicylic acid |
| (2) | Other nonsteroidal anti-inflammatory drugs |
| (3) | Sulfonamides |
| (4) | Anticoagulants |
| (5) | Pyrazolone derivatives (phenylbutazone, oxiphenbutazone, sulfipyrazone) |
| (6) | Monoamine oxidase inhibitors |
| (C) | Inhibitors of urinary excretion of sulfonylurea and active metabolites |
| (1) | Probenecid |
| (2) | Aspirin and salicylic acid |
| (3) | Other nonsteroidal anti-inflammatory drugs |
| (4) | Allopurinol |
| (5) | Sulfonamides |
| (D) | Augmentors of the effect of sulfonylurea |
| (1) | Alcohol |
| (2) | Aspirin and salicylic acid |
| (3) | Guanethidine and betanidine |
| (4) | Monoamine oxidase inhibitors |
| (E) | Agents that attenuate sulfonylurea action and may counteract the antihyperglycemic effect: |
| (A) | Enzyme inducers that increase sulfonylurea elimination |
| (1) | Alcohol (chronic, moderate use) |
| (2) | Barbiturates (chronic, moderate use) |
| (3) | Rifampicin |
| (B) | Agents that antagonize sulfonylurea action |
| (1) | B-Blockers |
| (C) | Inhibitors of insulin secretion or insulin action |
| (1) | Thiazides and loop diuretics |
| (2) | Diazoxide |
| (3) | B-Blockers |
| (4) | Phenytoin |
| (5) | Corticosteroids |
| (6) | Estrogens |
| (7) | Indomethacin |
| (8) | Isoniazid |
| (9) | Nicotinic acid |
| (D) | Mechanism obscure |
| (1) | Phenothiazines |
| (2) | Acetazolamide |
extent of a certain interaction may vary between the different sulfonylureas.

**Toxicity**

Most idiosyncratic toxic effects of sulfonylureas are said to appear within the first 2 months of treatment. They include the following:

- **Blood:** Agranulocytosis, thrombocytopenia, bone marrow aplasia, red cell aplasia, hemolytic anemia
- **Skin:** Rashes, pruritus, erythema nodosum, erythema multiforme, Stevens–Johnson syndrome, exfoliative dermatitis, purpura, photosensitivity
- **Gastrointestinal tract:** Nausea, vomiting, heartburn
- **Liver:** Abnormal function tests, jaundice, cholestasis, granulomatous hepatitis
- **Lung:** Possible diffuse pulmonary reaction
- **Thyroid:** Weak antithyroid activity
- **Kidney:** Antidiuresis, water retention (chlorpropamide)
- **Vasomotor:** Alcohol flush, tachycardia, headache (mainly chlorpropamide)
- **Cardiovascular:** Vasculitis

**Contraindications**

The following patients should not be treated with sulfonylureas:

1. Patients with type 1 diabetes or "pancreatic" diabetes
2. Pregnant women (only glyburide treatment is acceptable)
3. Patients with type 1 diabetes or "pancreatic" diabetes
4. Patients with a history of severe adverse reactions to sulfonylureas
5. Patients with severe infections, stress, or trauma
6. Patients with severe infections, stress, or trauma
7. Patients particularly prone to develop hypoglycemia, e.g., with liver or kidney disease.

**Conclusion**

Sulfonylureas and meglitinides are a class of glucose-independent insulin secretagogues which act by closing the pancreatic β-cell plasma membrane K\(_{ATP}\) channel. They are effective in improving glycemic control in patients with T2DM who are inadequately controlled on lifestyle modification. They are effective in mild to moderately hyperglycemic patients who still have modest pancreatic insulin reserve. They improve fasting plasma insulin levels. They do not correct the defect in the delay in early meal-mediated insulin secretion but increase late meal-mediated insulin secretion. Their effect on glycemic control has limited durability. Hypoglycemia is the major side effect and it can be the cause of significant morbidity and mortality. Modest weight gain occurs with sulfonylurea therapy. Cardiovascular safety during sulfonylurea therapy continues to be an unanswered question.

While the different sulfonylureas may have different intrinsic activity due to their binding characteristics for the K\(_{ATP}\) channel, the clinical benefits are the same when the appropriate dose is prescribed. Differences in hypoglycemia, dosing and duration of effects are due to differences in rates of absorption, plasma half-life, modes of metabolism, and methods of excretion.

Glucose-dependent insulin secretagogues are as or more effective in controlling glycemia and do not have the weight gain or hypoglycemia associated with sulfonylurea therapy. However, sulfonylureas are unique in being able to ameliorate the inability of the ATP/ADP ratio to facilitate insulin secretion and are the treatment of choice for such disturbances.

**References**

32 Billman G: The cardiac sarcolemmal ATP-sensitive potassium channel is a novel target for anti-arrhythmic therapy. *Pharmacology & Therapeutics* 2008;120:54–70.


170 Cambell IW: Comparing the actions of older and newer therapies on body weight: to what extent should these effects guide the selection of antidiabetic therapy. *International Journal of Clinical Practice* 2010;64:791–801.


CHAPTER 43

Metformin and other biguanides: pharmacology and therapeutic usage

Susan L. Samson and Alan J. Garber
Department of Medicine, Baylor College of Medicine, Houston, TX, USA

Key points

- Metformin is a first-line oral therapy for type 2 diabetes mellitus and is available in long-acting form and in combination formulations with other oral diabetes medications.
- The structure of metformin is \( N_1,N_1 \)-dimethyl biguanide, a derivative of guanidine.
- Metformin inhibits mitochondrial complex I resulting in altered AMP/ATP ratios in the cell. One consequence is activation of AMP-activated protein kinase which has pleiotropic effects on lipid and glucose metabolism.
- Metformin improves glycemic control mainly through inhibition of hepatic gluconeogenesis, and there also are improvements in peripheral glucose uptake and lipid metabolism.
- Metformin has positive effects in other metabolism-related diseases including polycystic ovary syndrome, nonalcoholic fatty liver disease, and HIV lipodystrophy.
- A common side effect is gastrointestinal upset which is usually self-limiting and rarely causes discontinuation of the medication.
- Although the risk of lactic acidosis with phenformin was significant, the risk of lactic acidosis with metformin approaches zero within appropriate prescribing practices.

Introduction

Metformin is the first-line therapy in the pharmacologic treatment of type 2 diabetes mellitus (T2DM), both as monotherapy and in combination with other medications. There are 40–50 years of experience with metformin in Canada and Europe, with a more recent introduction to the US market in 1995. The chemical structure of metformin is related to guanidine (Figure 43.1). The glucose-lowering properties of guanidine derivatives were recognized through the study of French lilac (\textit{Galega officinalis}, also known as Italian fitch, goat’s rue, Spanish sainfoin, professor-weed, and false indigo), which was used as a remedy for symptoms of glucosuria as early as medieval Europe (Figure 43.2). The high isoamylene guanidine (galegine) content of \textit{G. officinalis} was responsible for the hypoglycemic properties of this perennial herb [1].

Extracts from \textit{G. officinalis} were used to treat diabetes in the 1920s but had prohibitive gastrointestinal side effects [1]. The diguanides (two guanidine molecules connected by a carbon chain, Figure 43.1) were synthesized in the 1920s, but were hepatotoxic [1]. Biguanides (two linked guanidine molecules minus an ammonia radical) were synthesized in the 1950s and of the 300 or so biguanides synthesized, three have glucose-lowering properties: phenformin, buformin, and metformin. These drugs were used to treat diabetes for nearly 20 years but phenformin and buformin were withdrawn from the market in most countries by 1978 because of an association with fatal lactic acidosis [2]. Metformin has a very low rate of lactic acidosis compared with phenformin but was guilty by association. Additionally, in the late 1970s, the University Group Diabetes Program trial reported an increased risk for cardiovascular mortality with the prolonged use of oral hypoglycemic agents, including phenformin, compared with the use of insulin [3]. Although flaws in the design and conclusions of the trial later emerged, these events precluded the approval of metformin in the United States for nearly two decades.

Resurgence of interest in metformin occurred in the 1980s and 1990s based on clinical studies demonstrating its efficacy and safety in T2DM, especially with publication of the United Kingdom Prospective Diabetes Study (UKPDS) [4]. Metformin was shown to have efficacy similar to that of sulfonylureas in reducing fasting plasma glucose (FPG) and postprandial glucose concentrations, but did not cause weight gain or hypoglycemia in contrast to sulfonylurea therapy [4,5]. Pursuant to the UKPDS findings, and subsequent favorable recommendations regarding its use, metformin is the most commonly used diabetes drug throughout the world.
Figure 43.1 Chemical structures of guanidine derivatives used to treat hyperglycemia. The nonpolar hydrocarbon side chain of the biguanides (highlighted area) determines the degree of lipophilicity of the molecule and binding to the membrane.

**Pharmacology**

The unique hydrocarbon side chains of the biguanides (Figure 43.1) determine their lipophilicity and membrane binding affinity. The phenformin side chain contains a phenyl group, enabling it to permeate membranes including those of mitochondria. Metformin has a methyl side chain and exists as a hydrophilic cation or base at physiologic pH, precluding passive diffusion through cell membranes [6]. As a result, uptake of metformin into cells depends on organic cation transporters (OCT) of the SLC22A gene family. Metformin is absorbed in the intestine mediated by the plasma membrane monoamine transporter (PMAT), and perhaps aided by OCT3 on the luminal side of the enterocyte [7]. OCT1 and OCT3 proteins aid in the transport of metformin into the interstitial fluid on the basolateral side, and also are involved in uptake by the liver, skeletal muscle, and heart [6,7]. Metformin does not bind plasma proteins and is filtered unhindered at the glomerulus where renal tubule cells take up metformin mainly via OCT2 [6]. Unmodified metformin is excreted by the kidney by active tubular secretion mediated by the human multidrug and toxin extrusion 1 protein (MATE1), as well as MATE2K and OCT1 [6,7]. The pharmacokinetics of metformin varies among individuals explaining the variability seen in the glycemic responses to metformin. This may in part be due to genetic polymorphisms altering expression and/or function of the metformin transporters which alter absorption or renal clearance [6,7]. Cationic drugs, such as cimetidine, requiring the same transporters as metformin in the kidney can alter metformin clearance [6].
Mechanism of action

Mitochondrial respiration and AMP-activated protein kinase

Metformin has pleiotropic effects on glucose and lipid metabolism in clinical studies and researchers have begun to unravel the underlying mechanisms (Figure 43.3). Guanidine was reported to decrease mitochondrial oxidative phosphorylation in the 1950s [8]. Similarly, metformin has been demonstrated to inhibit complex I of the mitochondrial respiratory chain [9,10]. In simple terms, this would decrease entry of NADH into the respiratory chain and lower the proton motive force, decreasing ATP synthesis at complex V, increasing relative levels of AMP (Figure 43.3). Overall, the cell would perceive that there is an energy deficit and engage in metabolic processes to relieve that stress. Related to this, a key finding in the last decade or so is that metformin activates AMP-activated protein kinase (AMPK) [11].

AMPK is the energy sensor of the cell by detecting changes in the ratio of AMP or ADP to ATP. AMP binds the regulatory γ subunit to allosterically activate AMPK and promote phosphorylation at Threonine 172 (Thr172) of the catalytic α subunit [12]. AMP binding also protects AMPK dephosphorylation by cell phosphatases. AMPK phosphorylates key enzymes and transcription regulators [13]. The net effect is the downregulation of energy-consuming, anabolic processes, such as gluconeogenesis and lipid synthesis, along with activation of ATP-generating catabolic processes, such as fat oxidation (Figure 43.3). A key upstream AMPK kinase known to modify Thr172 is the Peutz-Jeghers tumor-suppressor gene product (LKB1) [12].

Figure 43.3 Schematic diagram of the biochemical effects of metformin in different cell types. Metformin inhibits mitochondrial complex I, reducing ATP production and increasing the AMP/ATP ratio in the cell. In the hepatocyte, AMP decreases the activity of fructose-1,6-bisphosphatase and the reduced ATP decreases the activity of pyruvate kinase to reduce gluconeogenesis. Additionally, the increased AMP/ATP ratios activate AMP-activated protein kinase (AMPK) which inhibits gluconeogenic gene expression. In the adipocyte, AMPK activation decreases hormone sensitive lipase activity, reducing lipolysis. Additionally, acetyl CoA carboxylases (ACC) 1 and 2 are inactivated by AMPK, resulting in decreased malonyl CoA production and increased transfer of fatty acids into the mitochondria via carnitine palmitoyltransferase (CPT). The net result is decreased fatty acid synthesis and increased beta oxidation. In skeletal muscle, AMPK increases glucose transporters at the plasma membrane to augment glucose uptake.
Although AMPK is thought to be a central player in the actions of metformin, the pathway from metformin to AMPK is not fully elucidated. Shaw et al. [14] demonstrated that loss of the upstream kinase LKB1 abrogated AMPK activation and the glucose lowering effects of metformin in vivo in a mouse knockout model. Thus, the downstream activation of AMPK by LKB1 appeared independent of fluctuations in cellular ATP content. However, the absolute requirement for LKB1 has been questioned by experiments demonstrating that metformin can dampen hepatocyte gluconeogenesis in the absence of LKB1, and even AMPK [15,16], due to a decrease in cellular energy state [15,16]. This comes full circle to an early hypothesis of energy charge caused by metformin, promoting pyruvate kinase activity, removing phosphoenolpyruvate from the gluconeogenic pathway because of its conversion to pyruvate. Therefore, a decrease in cellular energy charge caused by metformin could be the basis for acute changes in gluconeogenesis, with additional longer term effects involving downregulation of gluconeogenic enzyme expression as mediated by AMPK.

**Peripheral glucose uptake**

Although gluconeogenesis is recognized as the major metabolic process affected by metformin, clinical data show that metformin increases insulin-mediated whole-body glucose uptake [27] and glucose utilization 20–50% [28,29]. Skeletal muscle is responsible for approximately 80% of glucose uptake and utilization, so improvements in glucose disposal are expected to involve this important metabolic tissue. Euglycemic-hyperinsulinemic clamp studies in human subjects have been divided on the evidence for [29–32] and against [27,28,33,34] improved peripheral insulin sensitivity with metformin, perhaps due to subtle methodological differences.

**In vitro** studies have demonstrated that metformin can increase glucose uptake into muscle cells using rodent and human cell lines, primary culture, and isolated muscle tissue [35–39]. The rate of glucose uptake in skeletal muscle is dependent on facilitated diffusion via glucose transporters (GLUT) in the plasma membrane (PM). Insulin signaling induces GLUT4 translocation to the PM, while GLUT1 is already present in the PM at basal levels. From the experimental data, the observed increases in glucose uptake appear to be independent of, and additive to insulin treatment with the exception of one study [35]. On its own, metformin increases PM-associated levels of GLUT1 [40] and GLUT4 [41] in L6 myotubes. Notably, most in vitro experiments have been performed with what are supratherapeutic levels of metformin (~1–2 mmol L−1), which would not be achieved with in vivo treatment with metformin (μmol L−1 range), so caution is required in extrapolating these results to the clinical realm [42].

AMPK also may be central to the above observations since AMPK activation increases GLUT4 transcription and translocation [43,44] and contributes to metformin-induced GLUT4 translocation in muscle cell lines [37,41]. Reductions in GLUT4 endocytosis with chronic insulin stimulation also are mediated by AMPK in cardiomyocytes [45] and adipocytes [46].
Importantly, metformin has been shown to activate AMPK in vivo in human subjects with T2DM in both muscle [47] and adipose tissue [48], so that these in vitro observations may be relevant to the clinic.

**Lipid metabolism**

Metformin has well-documented favorable effects on serum lipid and lipoprotein parameters [49], with the most pronounced effect on serum triglycerides. In vivo, treatment of subjects with T2DM for 10 weeks with metformin versus a sulfonylurea results in increased phosphorylation of AMPK in adipose tissue biopsies [48]. In human primary adipocytes, metformin (at 2 mmol L⁻¹) inhibits β-adrenergic stimulation of lipolysis, dependent on AMPK [50]. Importantly, AMPK directly phosphorylates hormone-sensitive lipase, attenuating its activation by protein kinase A, decreasing lipolysis [48]. In vivo, this would decrease the amount of circulating free fatty acids (FFA), with less taken up by the liver, reducing hepatic secretion of very low-density lipoprotein (VLDL), and therefore hypertriglyceridemia [28,51].

Metformin also affects the interplay between fatty acid synthesis and β-oxidation through AMPK activation. AMPK phosphorylates and inactivates the acetyl CoA carboxylases (ACC) 1 and 2. The consequence is a decrease in malonyl CoA levels which discourages fatty acid synthesis while releasing carnitine palmitoyl transferase 1 (CPT1) from inhibition to facilitate entry of fatty acids into the mitochondria for β-oxidation [13,52]. The increased phosphorylation and inhibition of ACC by metformin has been reported in adipocytes [53], human adipose tissue [48], muscle [47], and liver [11]. It could seem paradoxical that a molecule implicated in inhibition of mitochondrial respiration would favor fatty acid oxidation. However, inhibition of complex I by metformin in vitro is rapidly reversed in the presence of high concentrations of FFA, promoting their β-oxidation [53]. AMPK also inhibits the transcriptional activity of sterol regulatory element-binding protein-1c (SREBP-1c), which regulates the expression of enzymes involved in de novo fatty acid synthesis, including fatty acid synthase [54]. Decreased synthesis and increased oxidation of fatty acids in the hepatocyte could be an explanation for decreased triglyceride storage in the liver seen with metformin treatment in animal models of obesity as well as some human studies [55,56].

**Glycemic control**

Metformin is the first-line pharmacologic treatment for patients with T2DM, supported by consensus statements by several professional diabetes associations throughout the world [57,58]. In early clinical trials, metformin monotherapy reduced fasting plasma glucose by approximately 50–90 mg dL⁻¹ (3–5 mmol L⁻¹) and A1c by 1.4–2.0% over a dose range of 1700 mg to 3000 mg daily [29,59,60]. Reductions in glycemic parameters are log-linear and dose-dependent allowing for incremental titration to effect [59]. Metformin is as effective as sulfonylureas or insulin in reducing FPG and A1c levels in patients with T2DM, but unlike sulfonylureas, fasting and postprandial serum insulin concentrations are unchanged or decreased with metformin [61,62]. Table 43.1 summarizes the accumulated metabolic data regarding metformin therapy derived from placebo- and active-controlled clinical trials in adults.

**Product formulations**

Metformin is available in immediate release (IR) and extended release (XR) tablets (Table 43.2). XR is equally efficacious in reducing A1c, fasting plasma glucose and postprandial glucose concentrations [63] and patients switched from IR to XR maintain comparable glycemic control. In healthy volunteers, maximum plasma concentrations achieved after a 2000 mg evening dose of XR are approximately 36% higher than following a 1000 mg evening dose of the IR formulation given twice daily [64]. Time to peak concentration is reached later (median, 7 h) with XR than with IR (approximately 2–3 h), but the extent of absorption, as measured by area-under-the-curve (AUC) of concentration by time, is similar between the formulations [64,65].

XR offers the advantage of once-daily administration at any dose and appears to have a better gastrointestinal tolerability profile [63]. Gastrointestinal adverse events were reported by 29% of patients receiving XR 1000 mg and by 39% of patients receiving IR 500 mg twice daily [66]. XR tablets contain a dual hydrophilic polymer matrix system which releases the drug slowly after ingestion. Fluid from the gastrointestinal tract enters the tablet and causes the polymers in the tablet to hydrate and swell [66]. Metformin is released through the process of diffusion and metered out over the dosing period. This process is independent of the pH of the gastrointestinal fluid and agitation.

In patients not achieving glycemic goals with metformin monotherapy, the addition of a second diabetes agent can produce significant and additive reductions in A1c and FPG. Also, with the multiple metabolic defects in T2DM, other agents...
can complement metformin, targeting additional physiologic defects. Combination of a second oral diabetes medication with metformin in the same tablet allows for simplified dosage regimens, particularly attractive among a population of patients that may have other comorbidities and require multiple medications. Metformin is available combined with glinides, sulfonylureas, thiazolidinediones, and dipeptidyl peptidase 4 inhibitors (Table 43.2). These agents will further lower the A1c (−0.64% to −0.97%) when added to metformin therapy [67]. The majority of combinations of these drugs with metformin are proven bioequivalent when administered in separate tablets or together in a single tablet. The combination glyburide/metformin is actually superior in clinical efficacy compared to equivalent doses given alone. In a randomized, double-blind crossover study, 2-h postprandial glucose excursion was significantly lower on day 14 after a mixed meal with the combination tablet than when given as the individual drugs [68]. The combination tablet has controlled particle sizes of glyburide, ranging from 6 to 21 μm so there is rapid release of glyburide at 2–3 hours, followed by a secondary peak 5–6 hours post dose [69]. In contrast, glyburide tablets administered separately with metformin tablets produce a single peak and twofold lower glyburide delivery in the first 3 hours [68].

The addition of metformin to insulin therapy in patients with T2DM further improves fasting glucose and A1c and can reduce insulin needs by approximately 25–36% in some studies, and in some cases obviating the need for insulin completely in other studies [70,71]. Additionally, when added to insulin therapy, metformin therapy does not induce hypoglycemia and can even promote weight loss [70,71].

**Table 43.1 Clinical effects of metformin**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glycemic control versus placebo</td>
<td>Fasting plasma glucose reduced by 3–5 mmol L⁻¹ (52–92 mg dL⁻¹)</td>
</tr>
<tr>
<td></td>
<td>A1c reduced 1.4–2.0%</td>
</tr>
<tr>
<td></td>
<td>Basic metabolic glucose clearance increased 30–53%</td>
</tr>
<tr>
<td>All-cause mortality</td>
<td>Reduced 36% compared to insulin or sulfonylurea therapy in overweight patients</td>
</tr>
<tr>
<td>Stroke</td>
<td>Reduced 36% compared to insulin or sulfonylurea therapy in overweight patients</td>
</tr>
<tr>
<td>Serum triglycerides</td>
<td>Reduced 15–45% dependent on baseline level</td>
</tr>
<tr>
<td>High-density lipoproteins cholesterol (HDL)</td>
<td>Unchanged or slightly increased</td>
</tr>
<tr>
<td>Low-density lipoproteins cholesterol (LDL)</td>
<td>Reduced 5–10%</td>
</tr>
<tr>
<td>Body weight</td>
<td>Variable with no loss reported in some studies, but other studies show loss in range of 0.7 kg to 5 kg with greater loss in heavier patients</td>
</tr>
<tr>
<td>Fibrinolysis</td>
<td>Reduces PAI-1 and tPA levels</td>
</tr>
</tbody>
</table>

The study enrolled 4075 patients with newly diagnosed T2DM who were randomly assigned to receive either “conventional therapy,” which was mainly dietary with a goal fasting glucose of less than 15 mmol L⁻¹ (270 mg dL⁻¹), or “intensive therapy” which involved active treatment with sulfonylurea, insulin, or metformin with a target fasting glucose of less than 6 mmol L⁻¹ (108 mg dL⁻¹). Over a mean of 10 years of follow-up, the treated patients had a median A1c of 7.0% compared with 7.9% in the conventional group [4]. The improvement in glycemic control with drug therapy was associated with a 12% risk reduction for any diabetes-related endpoint and a 25% risk reduction for microvascular complications compared to conventional therapy. It was more difficult to tease out the impact of diabetes treatment on macrovascular complications. However, a greater than expected benefit was seen for overweight patients on metformin. Although a somewhat smaller A1c difference was achieved on metformin therapy, there were significant reductions in risk for all-cause mortality and stroke versus intensive therapy with sulfonylureas or insulin [72]. A combined analysis of the main and supplementary UKPDS studies found a 19% reduction in risk in diabetes-related endpoints with metformin (p = 0.03). To date, metformin is the only antidiabetic agent to show an independent benefit on the macrovascular complications of diabetes, by reducing the risk for myocardial infarction. A supplemental study detracted from the positive findings with metformin because of data which showed a significant increase in diabetes-related and all-cause deaths with early addition of metformin to ongoing sulfonylurea therapy [72,73]. However, there were a small number of endpoints in the two groups (n = 269, 268), and additional epidemiologic analysis of the data did not confirm the negative effect of combined metformin and sulfonylurea therapy. Although the interpretation of cardiovascular data drawn from the UKPDS has been debated, additional smaller trials (cohort and randomized controlled)

**Microvascular and macrovascular risk reduction**

The efficacy of metformin in improving glycemic control and reducing microvascular complications of diabetes is similar to that of other oral agents used for the treatment of diabetes, related to A1c lowering [72]. The potential of metformin monotherapy to reduce adverse macrovascular endpoints in overweight T2DM patients first emerged in the UKPDS, which was undertaken to determine the long-term effects of improved glycemic control on complications of T2DM [5].
### Table 43.2 Metformin<sup>a,b</sup> product formulations

<table>
<thead>
<tr>
<th>Active ingredient(s)</th>
<th>Formulation</th>
<th>Strength (mg)</th>
<th>Dosing</th>
<th>Maximum daily dose (mg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metformin HCl</td>
<td>Immediate-release tablet</td>
<td>500, 850, 1000</td>
<td>500 or 850 mg three times per day with meals or ( \leq 2000 \text{ mg d}^{-1} ) can be given twice per day with meals</td>
<td>2550</td>
</tr>
<tr>
<td></td>
<td>Extended-release tablet</td>
<td>500, 750</td>
<td>Once per day with evening meal</td>
<td>2000</td>
</tr>
<tr>
<td></td>
<td>Liquid</td>
<td>100 mg mL(^{-1})</td>
<td>Twice per day with meals</td>
<td>2000</td>
</tr>
<tr>
<td>Diptediptyl peptidase IV inhibitor with metformin HCl</td>
<td>Sitagliptin/metformin tablet</td>
<td>50/500 50/1000</td>
<td>Twice per day with meals</td>
<td>100/2000</td>
</tr>
<tr>
<td></td>
<td>Sitagliptin/metformin extended-release tablet</td>
<td>50/500 50/1000</td>
<td>Once per day with evening meal</td>
<td>100/2000</td>
</tr>
<tr>
<td></td>
<td>Linagliptin/metformin tablet</td>
<td>2.5/500 2.5/850 2.5/1000</td>
<td>Twice per day with meals</td>
<td>5/2000</td>
</tr>
<tr>
<td></td>
<td>Saxagliptin/metformin&lt;sup&gt;c&lt;/sup&gt; extended release tablet</td>
<td>2.5/1000 5/500 5/1000</td>
<td>Once per day with evening meal</td>
<td>5/2000</td>
</tr>
<tr>
<td></td>
<td>Aloglipin/metformin</td>
<td>12.5/500 12.5/1000</td>
<td>Twice per day with meals</td>
<td>25/2000</td>
</tr>
<tr>
<td></td>
<td>Vildigliptin/metformin</td>
<td>50/500 50/850 50/1000</td>
<td>Twice per day with meals</td>
<td>100/2000</td>
</tr>
<tr>
<td>Insulin secretagogue with metformin HCl</td>
<td>Glyburide/metformin tablet</td>
<td>1.25/250 2.5/500 5/500</td>
<td>Twice per day with meals</td>
<td>20/2000</td>
</tr>
<tr>
<td></td>
<td>Glipizide/metformin tablet</td>
<td>2.5/250 2.5/500 5/1000</td>
<td>Twice per day with meals</td>
<td>20/2000</td>
</tr>
<tr>
<td></td>
<td>Repaglinide/metformin tablet</td>
<td>1/500 2/500</td>
<td>Twice per day with meals</td>
<td>4/2000</td>
</tr>
<tr>
<td>Thiazoladine-dione with metformin HCl</td>
<td>Pioglitazone&lt;sup&gt;d&lt;/sup&gt;e, metformin tablet</td>
<td>15/500 15/850</td>
<td>Twice or three times per day with meals</td>
<td>45/2550</td>
</tr>
<tr>
<td></td>
<td>Pioglitazone&lt;sup&gt;d&lt;/sup&gt;e, metformin extended release tablet</td>
<td>15/1000 30/1000</td>
<td>Once per day with evening meal</td>
<td>45/2000</td>
</tr>
<tr>
<td></td>
<td>Rosiglitazone&lt;sup&gt;d&lt;/sup&gt;f, metformin tablet</td>
<td>2/500 4/500</td>
<td>Twice per day with meals</td>
<td>8/2000</td>
</tr>
<tr>
<td></td>
<td>Rosiglitazone&lt;sup&gt;d&lt;/sup&gt;f, metformin extended release tablet</td>
<td>2/1000 4/1000</td>
<td>Once per day with evening meal</td>
<td>8/2000</td>
</tr>
</tbody>
</table>

<sup>a</sup>Metformin should not be used in patients with a serum creatinine \( \geq 1.5 \text{ mg dL}^{-1} (\geq 136 \text{ mmol L}^{-1}) \) in men and \( \geq 1.4 \text{ mg dL}^{-1} (\geq 124 \text{ mmol L}^{-1}) \) in women (USA) or creatinine clearance \( <60 \text{ mL min}^{-1} \) (Canada).

<sup>b</sup>Avoid in liver disease which is a risk factor for lactic acidosis.

<sup>c</sup>Maximum 2.5/1000 daily if on a Cyp3A4/5 inhibitor.

<sup>d</sup>Do not initiate with active liver disease or alanine transferase levels \( >2.5 \) times the upper limit of normal (ULN). Discontinue if alanine transferase levels remain \( >3 \) times the ULN or patient is jaundiced.

<sup>e</sup>Maximum 15 mg pioglitazone daily if on a strong Cyp2C8 inhibitor.

<sup>f</sup>Patients on rosiglitazone must be registered with the patient access program (USA) and sign informed consent (Canada).
also have reported a decrease in macrovascular complications with metformin [74–76].

The effects of metformin on macrovascular outcomes could be independent of glucose-lowering and the result of the influence of metformin on multiple cardiovascular risk factors [77], including serum lipid and lipoprotein parameters (Table 43.1), although the most pronounced effects are reductions in serum triglycerides [60,78]. Possible mechanisms are the decreased synthesis of very low-density lipoprotein (VLDL) particles, which are lowered with metformin therapy, inhibition of hepatic VLDL secretion, and increased clearance of chylomicrons [28,51,79]. Total and LDL-cholesterol concentrations are also reduced, although to a lesser extent. HDL-cholesterol concentrations are usually unchanged or slightly increased with metformin therapy.

Insulin resistance is accompanied by endothelial dysfunction, with pathologic disturbances of vasoregulatory and thrombosis-related substances secreted by endothelial cells. Plasminogen activator inhibitor 1 (PAI-1), which inhibits tissue plasminogen activator (tPA) and fibrinolysis, is elevated in insulin-resistant states. Using human adipose tissue in vitro, metformin (0.1–10 μM) decreases PAI-1 expression in a dose-dependent manner [80]. Clinically, metformin therapy reduces PAI-1 concentrations approximately 20% and the fibrinolytic response is enhanced [81,82]. Levels of tPA also are significantly reduced with metformin therapy [83,84].

Overall, metformin improves serum lipoprotein parameters, reduces body weight, and stimulates fibrinolysis, factors that potentially could result in reduced adverse macrovascular outcomes exclusive to metformin compared with other diabetes treatments. Further research is warranted to understand the true value and underlying mechanisms for metformin effects on cardiovascular health.

Additional therapeutic potential

Diabetes prevention

The Diabetes Prevention Program assessed whether intensive lifestyle modification or metformin therapy prevented or delayed the onset of T2DM in at-risk patients with impaired glucose tolerance (IGT) [85]. Over the 2.8 year study, the incidence of diabetes was reduced 58% with intensive lifestyle modification and 31% with metformin treatment compared with the placebo treatment [85]. There were no differences in outcomes by sex or ethnic group, but metformin was most effective in patients younger than 60 years of age and in individuals with higher body mass index (>35 kg m$^{-2}$) and higher baseline FPG [85]. Metformin therapy led to improvements in indices of insulin sensitivity and caused weight loss, which explained more than half of the effect on diabetes risk [86].

The Diabetes Prevention Program Outcomes Study (DPPOS) is an open-label extension of the DPP. The incidence of diabetes with metformin and lifestyle continued to be reduced compared to placebo up to 10 years after the original DPP randomization, but the metformin group had the lowest A1c and fasting glucose [87]. Reversion to normoglycemia just once during the DPP intervention, measured by FPG or glucose tolerance testing, reduced the risk of developing diabetes by more than 50% in the DPPOS extension [88], emphasizing the impact of early intervention in the prediabetic phase. Adherence was not the only factor affecting the response to metformin. Genetic analysis of the DPP metformin cohort revealed that the strongest genetic association with progression to diabetes was with polymorphisms in the AMPK γ subunit gene PRKAG2, with minor contributions by variations in the genes for the α1 and α2 subunits [89]. There also was an association with polymorphisms in cation transporters that mediate metformin transport into cells [89].

Metformin also has been shown to prevent progression to diabetes in patients with IGT in studies from India (−26%) and China (−77%) [90,91]. Metformin also prevents progression to diabetes by 50% in women with a history of gestational diabetes (GDM), compared to women with the same baseline glucose intolerance but without a history of GDM during pregnancy [92]. In spite of the data supporting the benefits of metformin for diabetes prevention, it currently is not approved by the U.S. Food and Drug Administration for diabetes prevention. Nor is it explicitly endorsed by guidelines from professional diabetes organizations at this time.

Diabetes in pregnancy

Tight glycemic control is warranted during pregnancy to reduce the risk for congenital abnormalities [93]. Metformin is afforded a pregnancy category B (FDA) rating with no evidence of teratogenicity in animal studies, but insufficient evidence of safety in human pregnancy. Transporters are present on the placenta facilitating exposure of the fetus to metformin [6]. Metformin is found in breast milk delivering less than 1% of the mother’s dose based on weight and does not cause hypoglycemia in the infant [93].

At this juncture, professional organizations do not explicitly endorse the use of metformin during pregnancy. However, there is data to support its use. In a large randomized trial, women with gestational diabetes were randomized to metformin or insulin [94]. Metformin use did not have adverse effects on neonatal outcomes, including prematurity and birth weight. Only 2% of subjects discontinued metformin due to gastrointestinal symptoms, and metformin was preferred compared to insulin, which caused a higher rate of severe hypoglycemia.

Weight loss

Metformin has been shown to reduce body weight independent of dosage. The mean weight loss is in the range of 1 to 5 kg, or a 1 to 3% reduction from baseline, reported in clinical trials in both diabetic patients and nondiabetic individuals [5,27]. Weight loss is greatest in the most obese patients [27] and is durable for at least 10 years in adherent patients from the
DPPOS [95]. Metformin also helps to counteract the weight gain caused by other oral agents including insulin, sulfonylureas, and thiazolidinediones [4,96]. Significant weight loss with metformin also has been achieved in women with polycystic ovary syndrome [97], patients with HIV lipodystrophy [98], and patients treated with weight-promoting antipsychotic and mood-stabilizing drugs [99]. Currently, weight loss is not an approved indication for use of metformin out of the context of treatment for diabetes.

**Polycystic ovary syndrome**

Polycystic ovary syndrome (PCOS) is a common endocrine disorder occurring among women of reproductive age and is the predominant cause of chronic anovulation and infertility.

The 2003 Rotterdam consensus broadened the diagnostic criteria of PCOS from the 1990 NIH consensus to include two of the following: (1) oligo- or amenorrhea, (2) clinical or biochemical signs of hyperandrogenism, and (3) polycystic ovaries by ultrasound [100]. Notably, being overweight or obese is not part of the criteria, but the prevalence of obesity in this population ranges from a low of 10% to 60–80%, dependent on ethnic background [100]. Insulin resistance is present in 70% of PCOS patients [101] and is compensated by hyperinsulinemia which has a central role in the pathogenesis of PCOS [102].

There is an insulin-induced increase in cytochrome P450c-17α activity in both the ovaries and adrenals [102]. This enzyme converts progesterone to 17α-progesterone through to androstenedione, which is then converted to testosterone, causing clinical manifestations of hyperandrogenism. The elevated insulin also inhibits sex hormone binding globulin production (SHBG) from the liver, further elevating free androgen levels. Liver production of insulin-like growth factor (IGF) binding protein 1 (IGFBP-1) is lowered by insulin, resulting in increased circulating free IGF-1 which binds its receptor on the ovarian thecal cell to synergize with luteinizing hormone (LH) to increase androgen production [103]. Insulin itself is able to cause cross-talk at the IGF-1 receptor further contributing to androgen production [102]. In a systematic review of 20 trials, metformin increased circulating SHBG and lowered total testosterone, dihydroepiandrosterone sulfate (DHEA), and androstenedione levels compared to pretreatment levels [104]. P450c17-α activity was decreased by metformin in both the ovaries and the adrenal glands, resulting in reduced production of 17-α-hydroxyprogesterone [105]. In spite of the positive changes in serum androgen levels, the effect of metformin on hirsutism often is not clinically significant for the patient, and adjunct antiandrogen therapy is required [100].

Neuroendocrine defects found in PCOS patients include an increase in gonadotropin-releasing hormone (GnRH) pulse frequency, with increased LH pulse amplitude and decreased follicle-stimulating hormone (FSH) secretion [106], but whether this is an independent defect, or is tied to insulin resistance has not been fully elucidated. Although insulin can augment GnRH-induced LH secretion from pituitary cells **in vitro**, insulin infusions do not influence gonadotropin levels including the response to GnRH [102].

Clinically, LH pulse amplitude in PCOS is reduced by metformin treatment accompanied by an improvement in the LH:FSH ratio [107], likely contributing to the improved menstrual regularity and ovulation seen in early studies with metformin treatment [108]. Some trials have reported that metformin alone is inferior to clomiphene in accomplishing conception, pregnancies, or live births in PCOS patients; nor does it seem that combination of metformin with clomiphene confers additional benefits [109]. However, a meta-analysis from the Cochrane collaboration using data from 16 placebo-controlled trials reported that metformin increased ovulation and clinical pregnancy rates, even when added to clomiphene, although live births were not increased [110]. Also, metformin may attenuate the ovarian response to gonadotropins for **in vitro** fertilization [111]. With these data in mind, the most recent consensus recommends metformin for PCOS patients only in the context of treatment of IGT or overt T2DM [100].

Women with PCOS have increased prevalence of IGT, metabolic syndrome, and T2DM compared to BMI-matched controls [112]. There are broad ranges for IGT and diabetes incidence when amalgamated from different studies—1.5 to 19% for IGT and 2.6 to 8.3% for T2DM—but this is undoubtedly much higher than for the general female population [112]. Conversion from IGT to diabetes also occurs in a shorter time frame [112], so there is a possible role for metformin to delay diabetes onset. The effects of metformin on weight loss and insulin sensitivity in PCOS have been inconsistent, with some studies showing a benefit and others not.

**Nonalcoholic fatty liver disease (NAFLD)**

Obesity and T2DM are significant risk factors for nonalcoholic fatty liver disease (NAFLD), which is now the most common liver disorder in the developed world [113]. NAFLD is present in one quarter to half of diabetes patients [113], and is characterized by increased triglyceride accumulation in hepatocytes (steatosis), with progression to chronic inflammation and necrosis (nonalcoholic steatohepatitis or NASH), and finally fibrosis and cirrhosis in about 20% of patients with NASH [114]. Hepatocellular carcinoma (HCC) is the third cause of cancer death and can form in the setting of NASH. Patients with T2DM have a two- to threefold increased risk of developing HCC [115], and this risk soars when combined with another factor, such as chronic viral hepatitis or alcoholism. Compared to insulin or sulfonylureas, metformin may prevent the development of HCC by up to 70% in diabetes patients [116]. What is not clear is whether the apparent risk reduction is due to an “anti-tumor” property of metformin, improvements in the metabolic derangements which lead to NAFLD, or both.

Metformin alters fatty acid metabolism in the hepatocyte mediated by AMPK, inhibiting de novo synthesis while increasing β-oxidation [11,58]. In experimental systems, this leads to decreased accumulation of triglycerides in the liver [55].
In several clinical studies, metformin administration to overweight or obese nondiabetic or diabetic patients for periods ranging from 4 to 12 months lowered liver transaminases [113]. One study reported a rise in enzyme levels back to pretreatment levels at 3 months of treatment [113]. Even so, in a subset of their patients, liver histology showed decreased inflammation, steatosis, and/or fibrosis at 12 months [156]. Seven studies (one randomized and six open-label design) have examined liver histology in adult patients, with four reporting decreased inflammation and fibrosis [113]. All told, in a meta-analysis, metformin was not found to improve histology, lobular inflammation, hepatocyte ballooning, or fibrosis [117]. Weight loss of ≥7% improves the histologic NAFLD activity score [117], so it is unclear if the heterogeneity seen among the trials is secondary to differences in weight changes in subjects.

**HIV lipodystrophy**

HIV-associated lipodystrophy is characterized by dyslipidemia and insulin resistance with increased central fat and decreased peripheral and facial fat [118]. It is often, but not always, associated with highly active antiretroviral therapy (HAART), and up to 80% of patients on these medications will develop features of lipodystrophy [118]. Open-label metformin (850 mg three times per day) for 2 months decreased weight, visceral fat, fasting insulin, and insulin levels during oral glucose tolerance testing in HIV patients on HAART [119]. A subsequent randomized controlled pilot study reported that 3 months of metformin therapy (500 mg twice per day) reduced weight, abdominal fat, and resulted in a 20% reduction in AUC insulin levels during an oral glucose tolerance test [120]. Extension of the study for 6 more months sustained the lowered insulin levels and caused further improvements in BMI and the cardiovascular risk marker tPA [121]. Importantly, lactate levels did not rise, given that patients on nucleoside reverse transcriptase inhibitors can have a higher risk for lactic acidosis [118]. A meta-analysis of six randomized controlled trials with 206 patients using metformin versus placebo also found decreased fasting insulin but without changes in glucose, triglycerides, LDL- or HDL-cholesterol [98]. BMI decreased but there was no effect on the waist-to-hip ratio [98]. The mean duration of treatment over all the studies was 27 months, but it is unclear if the positive effects are durable beyond that time, as HIV and lipodystrophy are chronic diseases. Also, there is no data available on effects on cardiovascular mortality with metformin in this population.

**Cancer biology**

From population studies, patients with diabetes have a higher cancer incidence than nondiabetics, including liver, pancreas, endometrium, colon, breast, and bladder cancer [122]. It is unclear if this is due to the existence of common risk factors for both diseases or whether hyperglycemia and/or hyperinsulinemia contribute to the increased risk by promoting cancer growth. Observational studies have hinted at a decrease in cancer incidence and mortality for diabetes patients treated with metformin including colorectal cancer, HCC, breast, and pancreatic cancer, and meta-analyses including these types of studies have reported somewhere in the range or 30–40% cancer incidence and mortality [123,124]. However, a systemic review including 14 randomized controlled trials, equivalent to 50,000 person-years, failed to find a significant effect of metformin on cancer incidence and all-cause mortality versus either usual care or other diabetes medications [125]. Nonetheless, this area clearly needs more clinical exploration, and awaits data from trials with metformin as an adjuvant to chemotherapy or prior to cancer surgery. Some of the positive effects of metformin in vivo could be through the lowering of glucose and insulin levels, and even circulating free IGF-1 levels which are elevated due to hyperinsulinemia [122]. Insulin can act as a mitogen, mediated through its activation of mitogen activated protein kinase (MAPK) and phosphoinositol 3-kinase/Akt pathways, which both regulate cell proliferation [126]. Akt activates mTOR (mammalian target of rapamycin) which also promotes glucose uptake into the cell [126], providing needed fuel to glycolytic pathways. The IGF-1 receptor is activated by IGF-1, or even insulin, and promotes cell proliferation as well [127].

Preclinical experiments have provided evidence that metformin can halt replication of cancer cell lines (including breast, ovarian, and prostate) in vitro independent of glucose and insulin [13]. One underlying mechanism may be through activation of the LKB1-AMPK pathway [13]. LKB1 was first studied in its role as a tumor suppressor before its association with AMPK was discovered. Knockdown of downstream AMPK expression attenuates the effect of metformin on cell proliferation, likely through releasing mTOR from inhibition. There are multiple additional signal transduction pathways in cancer cells that are affected by metformin in vitro [13], and it will to be exciting to see if these discoveries translate into improved clinical outcomes for patients with cancer.

**Adverse events, contraindications, and side effects**

**Lactic acidosis**

Humans produce 15–20 mmol kg\(^{-1}\) of lactate per day from glycolysis and deamination of alanine to be converted to pyruvate in the liver. With biguanide inhibition of the mitochondrial respiratory chain, there is an increase in glycolysis which can result in increased lactate production. Biguanides also decrease hepatic uptake and utilization of lactate for gluconeogenesis. These actions increase the potential for biguanides to cause lactic acidosis (LA), defined by decreased arterial blood pH (<7.35), with elevated lactate levels (>4–5 mEq L\(^{-1}\) or mmol L\(^{-1}\)) and an anion gap. LA is the most severe adverse effect of biguanide therapy, and can be fatal in up to 50% of cases [128].

From early epidemiologic data, there were an estimated to 40–64 cases of LA per 100,000 patient-years with
Metformin and other biguanides: pharmacology and therapeutic usage

phenformin [129], whereas most reports put the risk for metformin somewhere between 2 and 9 [130]. The difference is due to the lipophilic side chain of phenformin (Figure 43.1) which allows the drug to penetrate lipid and mitochondrial membranes and more potently inhibit NADH-dependent respiration compared with metformin [9]. Also, phenformin is partially metabolized in the liver through hydroxylation and nearly 1 in 10 people have a genetic polymorphism that slows hydroxylation, reducing drug clearance, and predisposing individuals to developing LA [131]. In contrast, metformin is excreted unchanged in the urine [7].

A systematic review from the Cochrane Collaboration included prospective controlled clinical trials and observational cohort studies in T2DM patients treated with metformin, alone or in combination with other medications. There were no cases of LA in 347 trials with 70,490 patient-years of treatment, even though over 40% of the prospective trials in the analysis did not exclude patients with impaired renal function [130]. The design of most clinical studies would include risk factors for LA as contraindications to metformin therapy in the exclusion criteria. Therefore, in the confines of appropriate prescribing practices, the risk of LA with metformin approaches zero [132]. Outside of clinical trials, the reality is that up to 50–70% of patients are prescribed metformin in spite of at least one contraindication, such as renal insufficiency, congestive heart failure, or liver dysfunction [133]. Nonetheless, the incidence of LA remains low. In a general healthcare database from the UK with over 50,000 patients, the incidence for LA was 3.3 per 100,000 patient-years for those on metformin and was even higher for sulfonylureas at 4.8 cases [134]. It is also important to note that patients with T2DM already have a greater risk of LA than nondiabetics, as high as 9.7 episodes per 100,000 patient-years in one study, without a further increase in the risk while on metformin [135].

Only rarely has LA occurred among patients receiving metformin without any other contributory causes or conditions and, in the vast majority of cases of LA with metformin therapy, patients had concurrent illnesses which could also cause LA and precipitate the reaction. In deliberate overdose, metformin-associated LA was noted in 16% of cases, so this drug certainly is capable of causing lactic acid on its own when present in supratherapeutic levels [136]. Thus, the goal is to avoid metformin accumulation.

Globally, the prescribing threshold for acceptable renal function for metformin therapy varies widely. The U.S. Food and Drug Administration still recommends against metformin use if the creatinine is >1.5 mg dL^{-1} in men and >1.4 mg dL^{-1} in women. Other countries have more permissive thresholds using estimated GFR (eGFR), with metformin to be used with caution at <60 mL min^{-1} per 1.73 m^{2} (Canada) or <45 mL min^{-1} per 1.73 m^{2} (UK, Australia). All seem to agree that metformin is contraindicated with eGFR <30 mL min^{-1} per 1.73 m^{2}, when the clearance of metformin drops precipitously [133].

Additional named contraindications to metformin therapy are congestive heart failure requiring pharmacologic treatment (potential for hypoxia), hepatic impairment (lactate accumulation), and acute or chronic metabolic acidosis. Alcohol may potentiate the effect of metformin on lactate metabolism. Metformin therapy should not be used in patients with evidence of hypoxemia, dehydration, or sepsis. Generally, guidelines and consensus statements do not recommend use of metformin in the inpatient setting.

Metformin itself is not nephrotoxic. However, iodinated contrast agents may temporarily decrease renal function, resulting in metformin accumulation and increasing the risk of LA [137]. There is wide variation in the recommendations from radiologic professional organizations in several countries regarding metformin and contrast administration [137]. There are no documented cases of LA in metformin users with normal baseline renal function having a single computed tomography scan and a usual dose of contrast [137]. Nonetheless, the general consensus is that, in patients with normal baseline renal function, metformin should be stopped at the time of contrast administration and for 48 hours thereafter, although some organizations specify that this is required only if greater than 100 mL of contrast is administered.

Other adverse effects

The most common side effects occurring with metformin therapy are gastrointestinal in nature, occurring in 20–30% of patients [65], including diarrhea, nausea, bloating and flatulence, cramping and abdominal pain (Table 43.3). A significant percentage (20–30%) of a metformin dose is found in the feces due to intolerable adverse events [59]. Gastrointestinal tolerability can be improved by initiating metformin therapy at low doses with slow titration upward, taking metformin doses with meals, and using metformin XR which has a lower frequency of gastrointestinal adverse events than metformin IR [6,69].

Up to 30% of patients will have decreased absorption of vitamin B12, with decreased levels detected as early as 4 months of therapy, and symptomatic deficiency developing after 5–10 years [60]. In some studies, decreased B12 levels are accompanied by decreased folate and elevated homocysteine [139]. Rarely, the B12 deficiency can be severe and manifest as megaloblastic anemia or peripheral neuropathy [140,141]. Because B12 deficiency is progressive over duration of metformin therapy, screening and appropriate supplementation is supported by some in the field [142], although the appropriate intervals for testing are not guided by consensus statements at this time.

Conclusion

Metformin is first line and central to T2DM treatment. Its primary target is liver glucose production, but it also has
recent findings may be independent of AMPK due to direct effects on mitochondrial respiration. Metformin is especially amenable to use as a combination therapy with other oral diabetes agents, allowing for targeting of multiple defects present in T2DM. There also is potential for metformin use in other diseases with metabolic derangements including diabetes in pregnancy, polycystic ovary syndrome, nonalcoholic fatty liver disease, HIV lipodystrophy, and even cancer. However, in most of these disorders, further clinical data is needed before metformin will be approved and accepted as a mainstream therapy.

**References**


53 González-Barroso MM, Anedda A, Gallardo-Vara E, et al.: Fatty acids revert the inhibition of respiration caused by the antidiabetic


65 Glucophage (metformin hydrochloride tablets) and Glucophage XR (metformin hydrochloride extended release tablets) prescribing information. Bristol-Myers Squibb Company; 2001.


86 Kitabchi AE, Temprosa M, Knowler WC, et al.: Role of insulin secretion and sensitivity in the evolution of type 2 diabetes in the
97 Morin-Papunen LC, Koivunen RM, Ruokonen A, Martikainen HK: Metformin therapy improves the menstrual pattern with minimal endocrine and metabolic effects in women with polycystic ovary syndrome. Fertility and Sterility 1998; 69(4):691–696.
105 Nestler JE, Jakubowicz DJ: Lean women with polycystic ovary syndrome respond to insulin reduction with decreases in ovarian P450c17 alpha activity and serum androgens. Journal of Clinical Endocrinology & Metabolism 1997; 82(12):4075–4079.


CHAPTER 44
PPAR agonists in the treatment of diabetes

Sunder Mudaliar and Robert R. Henry
Section of Diabetes/Metabolism, VA San Diego HealthCare System, San Diego, CA, USA

Key points
• Insulin resistance is a major pathophysiologic abnormality in type 2 diabetes and is intricately involved in the development of not only hyperglycemia, but also dyslipidemia, hypertension, and atherosclerosis.
• The PPARs are a family of nuclear receptors comprised of three subtypes designated PPARα, PPARγ, and PPARδ.
• The thiazolidinediones are potent insulin sensitizers and through their PPARγ agonist effects, they improve insulin sensitivity and glycemic control, and also have potentially beneficial effects on lipid metabolism, vascular tone, and endothelial function.
• Pioglitazone is the only PPARγ still generally available and has favorable effects on diabetes prevention, lipid parameters, and surrogate CV markers. In one study, it reduced the composite secondary outcome of all-cause mortality, nonfatal myocardial infarction, and stroke in patients with type 2 diabetes and high CVD risk.
• Dual PPARα/γ agonists are currently being studied and seek to combine the glucose-lowering/insulin-sensitizing PPARγ agonist properties of the thiazolidinediones with the lipid-modifying PPARγ properties of fibrates. The development of such dual agonists would help improve the management of both type 2 diabetes and diabetic dyslipidemia.

Introduction
Type 2 diabetes (T2DM) is a chronic disease characterized by hyperglycemia and numerous other metabolic abnormalities which lead to disabling micro- and macrovascular complications. It is one of the leading causes of premature cardiovascular morbidity and mortality and is estimated to affect more than 366 million people worldwide in 2011 and 552 million people by 2030 [1]. Three major pathophysiologic abnormalities contribute to the development of hyperglycemia in patients with T2DM: impaired insulin secretion, excessive hepatic glucose output, and insulin resistance in skeletal muscle, liver, and adipose tissue [2]. Insulin resistance is intricately involved in the development of not only hyperglycemia, but also dyslipidemia, hypertension, hypercoagulation, vasculopathy, and ultimately, premature atherosclerotic cardiovascular disease [3]. This cluster of metabolic abnormalities has been variously termed the “insulin-resistance syndrome” or the “metabolic syndrome” [3]. The thiazolidinediones (TZDs) are potent insulin sensitizers and through their PPARγ effects improve insulin sensitivity at target tissues—adipose tissue, skeletal muscle, and liver and improve glycemic control when used as monotherapy and in combination with other antihyperglycemic agents [4]. Additionally, the TZDs have beneficial effects on blood pressure, vascular tone, and endothelial function. However, these effects have not always translated into improved cardiovascular outcomes. Further, the use of these agents has been impacted by adverse effects reported including weight gain, bone loss, fluid retention, edema and precipitation of CHF as well as a possible association with bladder cancer.

Although three compounds in this class were approved for clinical use, only one agent pioglitazone (Actos®) is currently in common clinical use in the United States. The first agent in this class, troglitazone (Rezulin®) was marketed in the US from March 1997 until it was withdrawn in March 2000, due to risks of hepatotoxicity. Rosiglitazone (Avandia®) was approved in the US in May 1999, but in 2010, after the publication of epidemiologic evidence of possible greater incidence of cardiovascular events [5] the use of rosiglitazone was severely restricted in the US and it is currently not available in several countries. Recently, in June 2013, after an independent rejudication of the Rosiglitazone Evaluated for Cardiovascular Outcomes and Regulation of Glycemia in Diabetes (RECORD) trial, the FDA concluded that rosiglitazone does not increase the risk for adverse cardiovascular outcomes. However, the restrictions on its use have not been lifted. As regards pioglitazone, due to the association with bladder cancer, the drug has been withdrawn from clinical use in France and Germany. In the US, the Food and Drug Administration (FDA), after an ongoing safety review of pioglitazone, updated the drug
labels for pioglitazone-containing medicines to include safety information that the use of pioglitazone for more than 1 year may be associated with an increased risk of bladder cancer. The FDA also recommends that pioglitazone not be used in patients with active bladder cancer and be used with caution in patients with a prior history of bladder cancer. The FDA continues to evaluate data from an ongoing 10-year epidemiologic study; (http://www.fda.gov/Safety/MedWatch/SafetyInformation/SafetyAlertsforHumanMedicalProducts/ucm226257.htm). interim data from the 8-year follow-up of this study of 193,099 eligible men and women with diabetes mellitus suggest that in the cohort analysis, there is no overall statistically significant increased risk of bladder cancer among patients ever treated with pioglitazone; (http://general.takedapharm.com/Trial-Disclosure/01-03-TL-OPI-524-8-year-Interim-Report.pdf).

Despite all the above, the TZDs still have a role in the treatment of T2DM, since this is the only class of drugs that primarily targets insulin resistance. These drugs can be used in patients with renal impairment and there is a low incidence of hypoglycemia with these agents. Further, pioglitazone, the only agent in this class still readily available, also has favorable effects on diabetes prevention, lipid parameters, and in one study, it reduced the composite secondary outcome of all-cause mortality, nonfatal myocardial infarction, and stroke in patients with T2DM who have a high risk of macrovascular events.

In this chapter, we review the mechanism of action, pharmacology, and clinical use of the thiazolidinedione PPARγ agonist, pioglitazone and also provide a brief description of dual PPAR agonists still in clinical development.

**Mechanism of action**

The PPARs (peroxisome proliferator-activated receptors) are a family of nuclear receptors comprised of three subtypes designated PPARα, PPARγ, and PPARδ or β (Figure 44.1) [6]. Activation of PPARs leads to the formation of heterodimers with retinoid-X receptors (FXRs). These PPAR-FXR dimers bind to DNA-specific sequences called peroxisome proliferator response elements, which stimulate or dampen the transcription of target genes. PPAR isoforms vary in selectivity and sensitivity to ligands and recruit distinct co-activators/co-repressors, resulting in the regulation of different sets of genes [6]. PPARα regulates genes involved in fatty acid uptake and oxidation, inflammation, and vascular function. PPARβ regulates genes involved in fatty acid metabolism, inflammation, and macrophage lipid homeostasis. PPARγ receptors are found in key target tissues and are important regulators of adipocyte differentiation, hepatic and skeletal muscle glucose metabolism, lipid homeostasis, macrophage function, and vascular endothelial biology [4,6] (Figure 44.2).

The TZDs are highly selective and potent agonists for PPARγ and favorably influence the expression of several genes involved in glucose metabolism, including adiponectin, GLUT1, GLUT4, leptin, TNFα, Akt (protein kinase B), and hepatic glucokinase [4,6,7].

It is important to note that major aspects of PPARγ action are still unresolved. It is not known why synthetic activation of PPARγ produces dramatic antidiabetic effects when there is no general deficiency in PPARγ function in obesity/insulin-resistant states. Also, while antidiabetic effects of PPARγ agonist drugs correlate well with binding affinities, some compounds with poor agonist activities have potent antidiabetic effects. In a recent study, Choi et al. demonstrated that phosphorylation of PPARγ by the protein kinase Cdk5 does not activate/suppress general receptor transcriptional activity, but changes the expression of specific genes like adiponectin. Cdk5-mediated phosphorylation of PPARγ is linked to obesity induced in mice by high-fat feeding. Several antidiabetic PPARγ ligands, with or without classical agonist properties, directly inhibit the action of Cdk5 on PPARγ and restore a more normal, non-diabetic pattern of gene expression. In addition, inhibition of this PPARγ phosphorylation in humans by TZDs is closely associated with its antidiabetic effects (Figure 44.3). This study suggests a novel model for a Cdk5-PPARγ link in the pathogenesis of obesity/diabetes and for the therapeutic action of PPARγ ligands in these disorders [8]. In a follow-up study, Choi et al. described a novel synthetic compound—SR1664 with a unique mode of binding to PPARγ, lacking classical transcriptional agonism and able to block Cdk5-mediated phosphorylation in cultured adipocytes/insulin-resistant mice [9]. Moreover, SR1664 had potent antidiabetic activity without causing fluid retention and weight gain like the TZDs. Also, unlike TZDs, SR1664 did not interfere with bone formation in culture. These data illustrate that newer antidiabetes drugs can be developed by specifically targeting the Cdk5-mediated phosphorylation of PPARγ.

In addition to improving insulin sensitivity and glucose metabolism, PPARγ agonism may also mediate the anti-hypertensive and anti-atherosclerotic effects of the TZDs. Human smooth muscle cells (VSMCs) express mRNA and nuclear receptors for PPARγ 1 [10]. These receptors are upregulated during vascular injury and are present in early human atheroma.

---

**Table 44.1** Tissue expression profile and metabolic effects of the peroxisome proliferator-activated receptors (PPARs).

<table>
<thead>
<tr>
<th>PPAR isoform</th>
<th>Tissue expression profile</th>
<th>Metabolic effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPARα</td>
<td>Liver, kidney, adrenal, skeletal muscle, adipose tissue, vascular tissues</td>
<td>Vascular, anti-inflammatory, fat oxidation</td>
</tr>
<tr>
<td>PPARγ</td>
<td>Adipose tissues, skeletal muscle, heart, liver, kidney, gut, macrophages, vascular tissues, islet cells</td>
<td>Reverse cholesterol transport</td>
</tr>
<tr>
<td>PPARδ</td>
<td>Ubiquitous</td>
<td>Adipogenesis, fat redistribution, glucose metabolism, fat oxidation, vascular, anti-inflammatory</td>
</tr>
</tbody>
</table>

Figure 44.1. Tissue expression profile and metabolic effects of the peroxisome proliferator-activated receptors (PPARs).
Figure 44.2 Possible sites of sensitizing actions of the thiazolidinediones.

Figure 44.3 Classical and nonclassical pathways of PPARγ activation.

1 Activation of the PPARγ receptor by ligands like pioglitazone results in the classical PPARγ activation effects like improved insulin sensitivity and glucose lowering along with the undesirable side-effects of weight gain, edema, and bone effects.

2 Cdk5-mediated serine phosphorylation of PPARγ is linked to obesity induced in mice by high-fat feeding. Phosphorylation of PPARγ by the protein kinase Cdk5 does not activate/suppress general receptor transcriptional activity, but changes the expression of specific genes like adiponectin, leading to decreased insulin sensitivity.

3 Several antidiabetic PPARγ ligands, with or without classical PPARγ agonist properties, directly inhibit the phosphorylation of Cdk5 on PPARγ and restore a more normal, nondiabetic pattern of gene expression.
and precursor lesions. Pharmacologic activation of PPARγ with pioglitazone inhibits VSMCs proliferation and migration with the potential to limit restenosis and atherosclerosis [11].

Also of note, dual PPARα/γ agonists are currently in development and seek to combine the glucose-lowering/insulin-sensitizing PPARγ agonist properties of the TZDs and the lipid-modifying PPARγ properties of fibrates. The use of such dual agonists should improve the management of T2DM and diabetic dyslipidemia. Aleglitazar is a potent dual PPAR agonist with balanced affinity for the PPARα and PPARγ isoforms. The pattern of gene activation with aleglitazar is distinct from that achieved by the combination of fenofibrate and pioglitazone [12]. However, in July 2013, following the results of a regular safety review of the aleglitazar AleCardio phase III trial, the independent Data and Safety Monitoring Board (DSMB) recommended a halt to the trial due to safety signals and lack of efficacy. Based on this recommendation, the further development of aleglitazar has been stopped.

**Glycemic effects**

As potent insulin sensitizers, the TZDs have major effects on glucose metabolism that result in significantly lower blood glucose levels in insulin-resistant patients with T2DM (Table 44.1). Several studies in populations across the world, have documented the glucose-lowering effects of pioglitazone when used as monotherapy agent and in combination with sulfonylureas, metformin, dipeptidyl peptidase-4 (DPP-4) inhibitors, glucagon-like peptide-1 (GLP-1) receptor activators and insulin. In this context it must be pointed out that although pioglitazone is approved by the FDA in the United States for use as monotherapy and in combination with the sulfonylureas, metformin, and with insulin, in Europe, pioglitazone is not approved for use in combination with insulin.

**Pioglitazone monotherapy**

In a large multicenter, double-blind, placebo-controlled clinical trial, 408 patients with HbA1c >7.0% and FPG >140 mg dL\(^{-1}\) were randomized to pioglitazone monotherapy or placebo [13]. After 26 weeks, patients treated with 15, 30, or 45 mg pioglitazone had significant mean decreases in HbA1c (range −1.00 to −1.60% difference from placebo) and FPG (−39 to −65 mg dL\(^{-1}\) difference from placebo). The decreases in FPG were observed as early as the second week of therapy; maximal decreases occurred after 10–14 weeks and were maintained until the end of therapy. In addition, in all pioglitazone groups, there were significant mean decreases in triglycerides, increases in HDL-cholesterol (HDL-C), and only small changes in total cholesterol and LDL. Of interest, the subset of patients naïve to

| Table 44.1 Relative efficacy of pioglitazone |

<table>
<thead>
<tr>
<th>Reduction in fasting plasma glucose (mg dL(^{-1})) Mean change from control group</th>
<th>Reduction in HbA1c (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Monotherapy</strong></td>
<td></td>
</tr>
<tr>
<td>Pioglitazone 15 mg</td>
<td>39</td>
</tr>
<tr>
<td>Pioglitazone 30 mg</td>
<td>41–58</td>
</tr>
<tr>
<td>Pioglitazone 45 mg</td>
<td>65–68</td>
</tr>
<tr>
<td><strong>Combination with sulfonylureas</strong></td>
<td></td>
</tr>
<tr>
<td>Pioglitazone 15 mg</td>
<td>39</td>
</tr>
<tr>
<td>Pioglitazone 30 mg</td>
<td>58</td>
</tr>
<tr>
<td><strong>Combination with Metformin</strong></td>
<td></td>
</tr>
<tr>
<td>Pioglitazone 30 mg</td>
<td>38</td>
</tr>
<tr>
<td><strong>Combination with DPP-4 agents</strong></td>
<td></td>
</tr>
<tr>
<td>Pioglitazone 30/45 mg</td>
<td>15–20</td>
</tr>
<tr>
<td><strong>Combination with GLP-1 agonists (Exenatide 10 mcg BID)</strong></td>
<td></td>
</tr>
<tr>
<td>Pioglitazone 30/45 mg (or Rosiglitazone 4/8 mg) ± Metformin</td>
<td>30</td>
</tr>
<tr>
<td><strong>Combination with insulin</strong></td>
<td></td>
</tr>
<tr>
<td>Pioglitazone 15 mg</td>
<td>35</td>
</tr>
<tr>
<td>Pioglitazone 30 mg</td>
<td>40</td>
</tr>
</tbody>
</table>

Since the above data were not obtained from simultaneous trials, the comparative data is only a rough approximation of the relative effectiveness as stage, severity of hyperglycemia, and type of patients studied varied in the above studies.

Source: References [13,14,17–23].
oral antidiabetic therapy had greater improvements in HbA1c and FPG (difference from placebo of −2.55% and −80 mg dL$^{-1}$ for the 45 mg group) compared with previously treated patients. The overall adverse event profile of pioglitazone was similar to that of placebo. Similar results were reported in another 23-week, placebo-controlled monotherapy trial which demonstrated that compared with placebo, 30 mg daily of pioglitazone significantly reduced FPG by 58 mg dL$^{-1}$ and HbA1c by 1.37% [14]. In addition, there were significant improvements in measures of insulin sensitivity and β-cell function. In the studies above, pioglitazone monotherapy was associated with weight gain of 1.3 to 1.5 kg from baseline with the 15/30 mg dose and 2.8 kg with the 45 mg dose, compared to a decrease of about 1.5 kg with placebo.

In the context of monotherapy, it is noteworthy that there are few prospective clinical trials in patients with T2DM that have directly compared the efficacy of the various available antidiabetic agents on metabolic and clinical outcomes. The largest of these, the UKPDS [15] observed a progressive failure of the glucose-lowering therapies used (metformin, sulfonylureas, and insulin) to maintain glycemic control. In this study, the stepwise addition of antihyperglycemic agents to achieve glycemic goals resulted in substantial therapeutic overlap between the groups, making it difficult to analyze the effects of individual therapies. Importantly, the only insulin sensitizer available at the time of the UKPDS was metformin. Since the TZDs are more potent insulin sensitizers and also have beneficial effects on β-cell function, it would be expected that their use as monotherapy would provide more durable glucose control. This was demonstrated in the ADOPT study (A Diabetes Outcome Progression Trial) [16] with rosiglitazone. In this study, 4360 recently diagnosed, treatment-naïve patients with T2DM were randomized to rosiglitazone, metformin, or glyburide in a double-blind manner and treated for a median of 4 years. The cumulative incidence of monotherapy failure, defined as a confirmed level of fasting plasma glucose of more than 4 years. The cumulative incidence of monotherapy failure, defined as a confirmed level of fasting plasma glucose of more than 180 mg dL$^{-1}$, by Kaplan-Meier analysis at 5 years was 15% with rosiglitazone, 21% with metformin, and 34% with glyburide. This represented a significant risk reduction of 32% for rosiglitazone, as compared with metformin, and 63%, as compared with glyburide. So far no similar data is available with pioglitazone.

**Combination therapy with sulfonylureas**

The addition of pioglitazone to existing sulfonylurea therapy has been shown to significantly improve glycemic control in several large studies. In a large multicenter study, 560 patients on a stable regimen of a sulfonylurea and with HbA1c levels >8.0% [17] were randomly assigned to receive sulfonylurea plus either once-daily pioglitazone 15 or 30 mg or placebo. After 16 weeks, as compared with patients treated with placebo, those receiving pioglitazone 15 mg and 30 mg had significant decreases from baseline in HbA1c by 0.9% and 1.3%, respectively. FPG levels also decreased by 39 mg dL$^{-1}$ and 58 mg dL$^{-1}$ with pioglitazone 15 mg and 30 mg pioglitazone. Both pioglitazone groups had significant improvements in lipid levels and the drug was well tolerated with similar rates of adverse events in all groups. Of note, body weight increased by a mean of 1.9 kg and 2.9 kg in the 15 mg and 30 mg pioglitazone groups and this increase correlated with improvements in glycemic control. There was a decrease of 0.8 kg in the placebo group.

**Combination with metformin**

In obese type 2 diabetic patients inadequately controlled on metformin alone, the addition of a thiazolidinedione further improves glycemic control, insulin sensitivity, and β-cell function to a clinically important extent. In a multicenter study, 328 patients with T2DM who had suboptimal control on maximum doses of metformin [18] were randomized to either pioglitazone+metformin or placebo+metformin. After 16 weeks, the addition of pioglitazone 30 mg significantly decreased FPG levels by 38 mg dL$^{-1}$ and HbA1c by −0.83% compared with the placebo. Body weight increased by a mean of 0.95 kg with pioglitazone and decreased by 1.36 kg with placebo. One hundred and fifty-four of the study subjects entered an open-label phase where all patients were treated with pioglitazone 30 mg+metformin for 72 weeks and had mean decreases from baseline of 1.36% in HbA1c and 63 mg dL$^{-1}$ in FPG.

**Combination with DPP-4 inhibitor**

The combination of pioglitazone and DPP-4 inhibitors has been studied in those failing on pioglitazone+metformin/SU treatment. In these patients, large, multicenter, placebo-controlled studies have demonstrated reductions in HbA1c and FPG of −0.4–0.7% and 15–20 mg dL$^{-1}$, respectively, with no significant change in body weight compared to placebo [19].

**Combination with GLP-1 agonist**

A thiazolidinedione and GLP-1 agonist combination has also been demonstrated to provide effective glucose control. In a multicenter study, 233 patients with T2DM who were suboptimally controlled with thiazolidinedione treatment (pioglitazone/rosiglitazone with or without metformin) were randomized to exenatide or placebo [20]. At 16 weeks, compared to placebo, the addition of exenatide reduced HbA1c by −0.98% (from a baseline of 7.9%), and body weight by 1.51 kg (baseline 97 kg). As expected, there were more GI side-effects in the exenatide arm and hypoglycemia occurred in 11% versus 7% in the placebo group. Of note, although patients in this study experienced modest weight loss, the study was not designed to determine whether simultaneous initiation of exenatide and TZD therapies would prevent weight gain associated with the use of TZDs. Also, weight reduction occurred without the use of additional lifestyle interventions as part of the study protocol.

**Combination with insulin**

At the outset, it must be stated that the use of TZDs in patients with T2DM who are on insulin treatment is not approved in
Europe. In the US, pioglitazone is approved for use in combination with insulin. However, the FDA has cautioned that because combination thiazolidinedione and insulin therapy is associated with an increased incidence of pedal edema and a potentially greater propensity for patients to develop CHF, pioglitazone should be used prudently in patients with pre-existing pedal edema, especially in those who have evidence of milder degrees of heart failure (NYHA Class 1 and 2). In patients with NYHA Class 3 and 4 cardiac status, treatment with pioglitazone is not recommended.

The efficacy and safety of pioglitazone in insulin-treated patients with T2DM was evaluated in a large, multicenter study in which 566 patients with T2DM on stable insulin regimens for >30 days and HbA1c > 8.0% were randomized to receive either 15 or 30 mg pioglitazone, or placebo [21]. Per study protocol, the insulin dose remained unchanged, but could be decreased in response to hypoglycemia. At the end of double-blind treatment, relative to baseline, patients receiving pioglitazone (15 mg or 30 mg) showed statistically significant mean decreases HbA1c of −1.0 and −1.3%. Although the incidence of adverse events was similar in all treatment groups, the incidences of weight increase, hypoglycemia and edema were higher among patients receiving insulin+pioglitazone. In this study, pioglitazone also significantly improved HDL-C levels (mean increases ranging in which 566 patients with T2DM were evaluated in a large, multicenter study in which 566 patients with T2DM on stable insulin regimens for >30 days and HbA1c > 8.0% were randomized to receive either 15 or 30 mg pioglitazone, or placebo [21]. Per study protocol, the insulin dose remained unchanged, but could be decreased in response to hypoglycemia. At the end of double-blind treatment, relative to baseline, patients receiving pioglitazone (15 mg or 30 mg) showed statistically significant mean decreases HbA1c of −1.0 and −1.3%. Although the incidence of adverse events was similar in all treatment groups, the incidences of weight increase, hypoglycemia and edema were higher among patients receiving insulin+pioglitazone. In this study, pioglitazone also significantly improved HDL-C levels (mean increases ranging from +7.1% to +9.3%) compared with baseline or placebo. The 30 mg dose also significantly reduced mean triglyceride levels (−23.7%) compared with placebo. No consistent changes in TC or LDL-C levels were observed.

A recent meta-analysis of eight trials examined the benefits of adding pioglitazone to an insulin regimen in 3092 patients with T2DM. It was found that pioglitazone doses of up to 45 mg d−1 were associated with a mean reduction in HbA1c of 0.58% and an average weight gain of 3 kg along with slightly more frequent hypoglycemia (relative risk 1.27; p = 0.06). Pioglitazone use was also associated with significantly less insulin requirements (weighted mean difference −0.19 U kg−1 d−1 or −12.03 U d−1). Peripheral edema was more frequent in the pioglitazone groups and where reported, HDL-C tended to increase with pioglitazone. None of the studies reported on fractures in women, and data on cardiovascular events were inconclusive, with most studies being too short or too small to assess these long-term outcomes [22].

**Metabolic effects**

As potent insulin sensitizers, the PPARγ agonists not only improve glycemic control, but also have multiple metabolic effects on peripheral insulin sensitivity, β-cell function, diabetes prevention, lipid metabolism, and body fat distribution [4].

**Insulin sensitivity**

Several studies have documented the *in vivo* effects of TZDs on peripheral insulin action in patients with T2DM, and in other insulin-resistant states like impaired glucose tolerance (IGT), polycystic ovary disease, previous gestational diabetes, and lipodystrophy [4]. Some of these studies have used the hyperinsulinemic-euglycemic clamp study which is currently considered the “gold standard” to evaluate peripheral insulin sensitivity, while others have used the frequently sampled intravenous glucose tolerance test (FSIVGTT) and other less direct methods, including the insulin tolerance test, the intravenous or oral glucose tolerance test and the HOMA-S to assess thiazolidinedione effects on insulin sensitivity. In all these studies, the TZDs have improved insulin-mediated peripheral glucose utilization in obese, insulin-resistant subjects with T2DM by approximately 30–100% [4].

In an elegant study using the two-step (40 and 160 mU min−1 per m2) hyperinsulinemic clamp, indirect calorimetry and hepatic tracer technology, Miyazaki et al. evaluated hepatic and peripheral insulin sensitivity in 23 obese diabetic patients on a stable dose of sulfonylurea [23]. After 16 weeks, the addition of pioglitazone 45 mg was associated with significant decreases in FPG (184 to 135 mg dL−1), HbA1c (8.9 to 7.2 %) and plasma FFA levels. Endogenous glucose production (EGP) during the first insulin clamp step was significantly decreased after pioglitazone treatment, whereas insulin-stimulated total and nonoxidative glucose disposal during the second insulin clamp was increased by 38% (p < 0.01). The change in FPG was related to the change in basal EGR, EGP during the first insulin clamp step, and total glucose disposal during the second insulin clamp step. These results suggest that pioglitazone therapy in type 2 diabetic patients decreases plasma glucose by improving hepatic and peripheral (muscle) tissue sensitivity to insulin.

In a large, international study, Pavo et al. studied the effect of 32 weeks of monotherapy with either pioglitazone or metformin on glycemic control and insulin sensitivity in 205 recently diagnosed type 2 diabetes patients who were naïve to anti-diabetic therapy [24]. In this study, pioglitazone was comparable to metformin in improving glycemic control as measured by HbA1c and FPG (decreases of ~1.3% and ~54 mg dL−1). However, insulin sensitivity (as assessed by HOMA-S) increased significantly by 17.4% in the pioglitazone group compared with a nonsignificant increase of 8.9% in the metformin group.

**Beta-cell function**

Along with defects in action in insulin sensitivity, defects in insulin secretion from the pancreatic β cells also play a major role in the pathogenesis of T2DM. Data from the UKPDS and other studies have suggested that β-cell function decreases with duration of T2DM whether treated with diet therapy or antihyperglycemic therapy [25,26]. Since PPARγ is expressed in human islet cells, both at the mRNA and protein levels, it is possible that the TZDs have direct effects on human pancreatic endocrine cells [27]. At present however, it is not possible to image β-cell mass *in vivo* in humans with sufficient accuracy and specificity. The best available methods of β-cell function *in vivo* are measurements of serum insulin levels at baseline and
after stimulation with glucose and arginine. Other methods include the homeostatic model assessment (HOMA), the intravenous glucose tolerance test, the hyperglycemic clamp and the pro-insulin to insulin ratio.

In one of the early studies of β-cell function [28], Miyazaki et al. found an increase in plasma insulin response during OGTT in diabetic patients treated with 30 and 45 mg d−1 of pioglitazone. An increased or maintained plasma insulin response in the presence of a decreased plasma glucose concentration suggests an improvement in β-cell function. The variable effect of TZDs on the plasma insulin response to a glucose challenge may be explained by two opposing effects. Firstly, a decrease in fasting and postprandial glucose concentrations leads to reduced “glucose toxicity” on the β cell and improved insulin secretion, and secondly, with the improvement in insulin sensitivity, there is a reduced requirement for insulin secretion.

In another study from the UK, 30 subjects with diet-controlled T2DM were randomized to treatment with pioglitazone or placebo. After 3 months, in the pioglitazone group, along with significant improvement in glycemic control, insulin sensitivity and the pro-insulin to insulin ratio, there was a significant improvement in HOMA-S: +11.5% versus −2.0% with placebo. However, there was no change in stimulated β-cell function as determined by hyperglycemic clamp test. As expected, there was a significant reduction in the HbA1c by 0.6% with pioglitazone compared with placebo. There was no significant weight gain associated with pioglitazone therapy: +0.7 kg versus +1.1 kg in the placebo group [29].

From the studies above, it is clear that the TZDs as PPARγ agonists have only modest β-cell effects in humans and reduced “glucose toxicity” alone may account for these effects. In contrast, in a mouse model, pioglitazone treatment had both an acute effect to preserve β-cell mass in diabetic mice directly by acceleration of cell differentiation/proliferation and suppression of apoptosis; and a chronic indirect effect through amelioration of the underlying metabolic disorder and deceleration of oxidative stress [30].

**Diabetes prevention**

It is well recognized that subjects with impaired glucose tolerance (IGT; plasma glucose between 140 and 199 mg dL−1 2 hours after a 75 gm oral glucose load) and impaired fasting glucose (IFG; plasma glucose between 100 and 125 mg dL−1) are not only insulin resistant, but also at high risk to progress to T2DM. Since the TZDs are potent insulin sensitizers, they have the potential to prevent T2DM in high-risk populations with prediabetes. In the landmark US Diabetes Prevention Program [31], among individuals with IGT and high FPG (95–125 mg dL−1), troglitazone use was associated with a robust 75% reduction in the incidence of diabetes compared to the placebo group [32]. In the same period (mean 0.9 years), there was a 57% reduction in the intensive lifestyle group and a 44% reduction in the metformin group. Of note, after troglitazone withdrawal due to liver toxicity, the diabetes incidence rate increased to that of the placebo group, thereby indicating that the protective effect did not persist after the drug was stopped. In the DREAM study [33], rosiglitazone at 8 mg daily for 3 years substantially reduced incident T2DM by 60% and also increased the likelihood of regression to normoglycemia in adults with IFG, IGT, or both. There was more heart failure in the rosiglitazone group. Evidence for diabetes prevention with pioglitazone is available from the ACT Now study in which 602 patients were randomized to pioglitazone or placebo [34]. After a median follow-up of 2.4 years, as compared with placebo, pioglitazone reduced the risk of conversion from IGT to T2DM by 72%. In addition, significantly more people converted to normal glucose tolerance with pioglitazone 48% versus 28% in the placebo group. Pioglitazone therapy was also associated with significant decreases in diastolic blood pressure (2.0 mmHg vs. 0.0 mmHg); reduced rate of carotid intima-media thickening (31.5%), and a greater increase in the level of HDL-C (7.35 mg dL−1 vs. 4.5 mg dL−1). As expected, weight gain was greater with pioglitazone than with placebo (3.9 kg vs. 0.77 kg) and edema was more frequent (12.9% vs. 6.4%). Although the evidence for diabetes prevention with the TZDs is strong, one has to keep in mind the risks for undue weight gain, edema, precipitation of CHF, bone fractures, and association with bladder cancer in some subjects. However, in selected subjects in whom the benefits outweigh the risks, pioglitazone use for diabetes prevention remains a potential option.

**Lipids**

In T2DM, the major quantitative change in lipid levels is an elevation in triglyceride-rich lipoproteins, a decrease in HDL-C concentrations and LDL-C levels that are often quantitatively similar to that in the general population [3]. However, qualitative changes in the composition of the LDL molecule, including an increase in the proportion of small, dense LDL which is prone to glycation and oxidation, tend to make it more atherogenic [3]. This diabetic dyslipidemic profile is closely related to underlying insulin resistance and may, in part, be responsible for the increased cardiovascular morbidity and mortality in type 2 diabetic patients. By improving glucose tolerance and reducing insulin resistance, the TZDs have the potential to favorably influence diabetic dyslipidemia. Treatment with the TZDs increases both HDL-C and LDL-C with variable effects on triglycerides. There are also differences between the TZDs. This was demonstrated in a double-blind, head-to-head study in 802 subjects with T2DM (diet/monotherapy treated) and dyslipidemia (treatment naive) who were randomized to either pioglitazone or rosiglitazone. After 24 weeks, despite equivalent glycemic control with both agents, there was greater lowering of triglyceride levels by pioglitazone (−51.9 mg dL−1 vs. +13.1 mg dL−1 with rosiglitazone; p < 0.001) [35]. Additionally, the increase in HDL-C was significantly greater (5.2 vs. 2.4 mg dL−1) and the increase in LDL-C was significantly less (12.3 vs. 21.3 mg dL−1) for pioglitazone compared with rosiglitazone. LDL particle concentration was also reduced.
with pioglitazone and increased with rosiglitazone ($p < 0.001$), while LDL particle size increased more with pioglitazone ($p < 0.005$). Thus in this study, pioglitazone and rosiglitazone had significantly different effects on plasma lipids independent of glycemic control or concomitant lipid-lowering treatment.

It is possible that these differential effects on lipids may account for the different cardiovascular outcomes with these agents. While rosiglitazone has not been associated with improvements in cardiovascular morbidity, pioglitazone may have cardiovascular benefits. In the large PROactive study, 5238 patients with T2DM who had evidence of macrovascular disease were randomized to pioglitazone or placebo [36]. In this study, as expected, LDL-C significantly increased by 7.2% with pioglitazone versus 4.9% with placebo. Although this increase is potentially detrimental, LDL subparticle analyses revealed a marked shift in the LDL particle size, resulting in more of the less atherogenic, buoyant Type A LDL particle and a decrease in the pro-atherogenic, small, dense Type B LDL particle levels [36]. Although the primary endpoint did not meet statistical significance, the main composite secondary endpoint of all-cause mortality, nonfatal myocardial infarction, and stroke was significantly lower by 16% in the pioglitazone group.

**Adipose tissue**

One of the intriguing aspects of treatment with the TZDs has been the observation that patients treated with the TZDs significantly improve their insulin sensitivity and glucose and lipid profiles, they tend to gain weight and accumulate adipose tissue. Considerable research has focused on the sites and nature of this adipose tissue weight gain. It is well known that the TZDs, through PPARγ activation cause preadipocytes to differentiate into mature fat cells and also cause the induction of key enzymes involved in lipogenesis [37]. At the tissue level, *in vitro* studies have demonstrated that the TZDs specifically promote the differentiation of pre-adipocytes into adipocytes in subcutaneous fat and not in omental fat [38]. Site differences in adipose tissue function appear to have implications for insulin-resistant conditions. In mature adipose tissue, subcutaneous adipocytes have higher leptin secretion, similar tumor necrosis factor-α secretion, and lower catecholamine-stimulated lipolysis as compared with omental adipocytes [39]. Thus it would appear that the accumulation of subcutaneous fat rather than visceral (omental) fat may be beneficial in terms of promoting insulin sensitivity.

*In vivo* studies in humans confirm that TZD treatment produces a shift in adipose tissue distribution from omental to the subcutaneous compartment. Determinations of fat distribution after pioglitazone therapy demonstrate a greater accumulation in the subcutaneous adipose tissue compartment at the expense of visceral fat. Miyazaki and coworkers [40] examined the effect of pioglitazone on abdominal fat distribution in 13 type 2 diabetic patients who were being treated with a stable dose of sulfonylurea ($n = 7$) or with diet alone ($n = 6$) and received pioglitazone 45 mg d$^{-1}$ for 16 weeks. Before and after pioglitazone treatment, subjects underwent a 75 g oral glucose tolerance test (OGTT) and an insulin clamp. Abdominal fat distribution was evaluated using magnetic resonance imaging at L4-5. After 16 weeks of pioglitazone treatment, there were significant improvements in FPG (144 to 131 cm$^2$; $p < 0.05$); and the ratio of visceral to SQ fat decreased from 0.59 to 0.44; $p < 0.01$). In the postabsorptive state hepatic insulin resistance correlated positively with visceral fat area ($r = 0.55; p < 0.01$) and the glucose metabolic clearance rates (MCRs) during the insulin clamp steps were negatively correlated with the visceral fat area [40]. These results demonstrate that a shift in fat distribution from visceral to subcutaneous adipose depots after pioglitazone treatment is associated with improvements in hepatic and peripheral tissue sensitivity to insulin. In another recent study in obese patients with T2DM with poor glycemic control [41], it was noted that intensification of insulin therapy alone for 3–4 months caused modest weight gain (1.7 kg) and no change in body fat distribution, lean body mass (LBDM), or liver fat. In contrast, the addition of pioglitazone, at equivalent glycemia, caused greater weight gain (4.9 kg) and increases in body fat mass mainly in the subcutaneous adipose tissue compartment with a trend towards a decrease in visceral adipose tissue and liver fat and an increase in LBDM. The authors concluded that these changes in fat distribution may contribute to the beneficial effects of pioglitazone, despite greater weight gain.

**Cardiovascular effects**

Ultimately, up to 80% of type 2 diabetics die from macrovascular cardiovascular disease. An added advantage for any diabetes drug is a beneficial effect to improve cardiovascular risk factors. Pioglitazone has multiple beneficial effects on the cardiovascular system and these are detailed below and in Table 44.2.

**Myocardial function and blood flow**

Myocardial blood flow is an important determinant of LV function. In a recent double-blind, placebo-controlled study, 26 patients with familial combined hyperlipidemia treated with conventional lipid-lowering therapy, were randomized to pioglitazone or placebo. Myocardial glucose uptake (MGU) and myocardial blood flow (MBF) were measured by positron emission tomography at rest and during adenosine-induced hyperemia during a euglycemic hyperinsulinemic clamp at baseline and after treatment. After 12 weeks, along with a favorable effect on blood lipid parameters, the addition of pioglitazone led to significant improvements in MGU and MBF [42].
Table 44.2 Cardiovascular risk factor reduction with pioglitazone treatment

- Improve dyslipidemia (↑ HDL-C, ↓ TG, ↓ LDL density)
- Decrease microalbuminuria, blood pressure
- Reduce carotid intimal medial thickness
- Reduce neointimal/VSMC (vascular smooth muscle cell) proliferation, macrophage migration
- Improve vascular reactivity, endothelial function
- Decrease vascular inflammation, CRP, Endothelin-1, MMP-9, MCP-1 levels
- Increase thrombolysis, decrease PAI-1 activity
- Increase adiponectin levels

Beneficial effects of pioglitazone on left ventricular (LV) function were seen in a study in 30 nondiabetic, insulin-resistant hypertensive patients. In these subjects, pioglitazone treatment for 6 months increased serum adiponectin levels and insulin sensitivity and also significantly improved LV diastolic function without changes in LV mass [43]. Similar improvements in LV function were seen in another recent study in 15 patients with T2DM where 6 months of pioglitazone treatment improved LV diastolic function and suppressed the synthesis of type III collagen (a marker of myocardial fibrosis). Discontinuation of pioglitazone weakened the suppression of the synthesis of type III collagen and was associated with worse LV diastolic function [44].

Notwithstanding the favorable effects on myocardial blood flow and function mentioned earlier, it must be noted that the use of the TZDs is associated with a propensity towards increased fluid retention and precipitation of congestive heart failure in some patients.

Blood pressure

The prevalence of hypertension in diabetic patients is up to twofold higher than in nondiabetic individuals and in T2DM and other insulin-resistant states there is impaired insulin-induced vasodilatation [45]. It is possible that by improving insulin action, the TZDs enhance the tonic vasodilator response to insulin and thereby reduce peripheral vascular resistance and blood pressure. Further, by reducing hyperinsulinemia and plasma insulin levels, it is possible that the TZDs attenuate the potential blood pressure raising actions of insulin, such as renal sodium retention and increased sympathetic activity [45]. In one study, 60 nondiabetic patients with arterial hypertension were randomized to pioglitazone 45 mg daily or placebo. After 16 weeks, along with significant improvements in insulin sensitivity and lipid parameters, SBP was significantly lower by 6.7 mmHg and DBP by 7.9 mmHg as compared to a decrease of 3.1 and 1.8 mmHg with placebo [46]. In the large PROActive study [47] with 5238 patients, blood pressure over 3 years was reduced significantly more in the pioglitazone-treated group than in the placebo group (median change in SBP 3 mmHg vs. 0 mmHg). Thus although small, the effects of the TZDs on blood pressure are consistent and may be beneficial in the long-term amelioration of cardiovascular complications.

### Glomerular function and albuminuria

Microalbuminuria (urinary albumin excretion rate between 30 and 300 mg per 24 h) is widely considered to be a marker of impaired vascular integrity in type 2 diabetic patients and an early indicator of renal and cardiovascular disease risk as well as of an increased risk of all-cause mortality. Studies across populations have demonstrated the ability of the TZDs to reduce urinary albumin excretion. In an early study from Japan [48], 45 type 2 diabetes patients with microalbuminuria were randomly assigned to treatment with pioglitazone 30 mg d⁻¹, glibenclamide 5 mg d⁻¹, or voglibose (an alpha-glucosidase inhibitor) 0.6 mg d⁻¹. After 3 months, only pioglitazone 30 mg daily was effective in reducing urinary albumin excretion and urinary endothelial-1 (ET-1) concentrations. Urinary ET-1 excretion is present in type 2 diabetes patients with microalbuminuria, and an increase in circulating ET-1 precedes the microalbuminuric phase of renal injury related to diabetes. Of note, pioglitazone reduces albuminuria in type 2 diabetic patients with hypertension and microalbuminuria even in the setting of concurrent treatment with renin-angiotensin system inhibitors [49].

In a recent large meta-analysis, treatment with TZDs as compared to placebo and active comparator, was found to significantly decrease urinary albumin and protein excretion in patients with diabetes [50]. Whether this effect is mediated through blood pressure reduction or via direct vasculoprotective mechanisms is not clear at present. Regardless, the TZDs improve blood glucose levels and reduce blood pressure and albumin excretion in patients with T2DM. However, so far, there is no evidence that these short-term benefits translate into favorable effects on hard renal/CVD outcomes in the long term.

### Anti-atherogenic effects

Atherosclerosis is a progressive vascular disease initiated by the accumulation of LDL-C in the subintimal space of the vessels and culminating in plaque rupture with resultant thrombosis and acute occlusion leading to a myocardial infarction or stroke. Three key components play a crucial role in the development of accelerated atherosclerosis, the endothelial cells (ECs), the monocyte/macrophage cells and the vascular smooth muscle cells (VSMCs) [51]. All of these major cells express PPARγ and there is a strong evidence from in vitro animal and human studies that pioglitazone as a PPARγ agonist has beneficial effects on the endothelium, VSMCs and macrophage function [51,52]. Pioglitazone increases the expression in macrophages of LXRx and ABCA1 (which facilitate the efflux of cholesterol from macrophages and reduces foam cell formation); inhibits VSMC cell proliferation and migration; and improves peripheral and coronary vascular endothelial function in patients with T2DM and CVD.

Thus, the PPAR agonists through their anti-atherogenic effects have the potential to slow the development of/improve the stability of the atherosclerotic plaque. So far the effects of these changes upon biochemical markers and surrogate markers of atherosclerosis progression (like carotid IMT) appear to be
favorable [53,54]. In the CHICAGO study [53], 462 patients with T2DM were randomized to pioglitazone (15–45 mg d\(^{-1}\)) or glimepiride (1–4 mg d\(^{-1}\)) as an active comparator. At week 72, the primary endpoint of progression of mean CIMT was significantly less with pioglitazone versus glimepiride (−0.001 mm vs. +0.012 mm, respectively; difference, −0.013 mm). Pioglitazone also significantly slowed progression of maximum CIMT compared with glimepiride (0.002 mm vs. 0.026 mm, respectively, at 72 weeks; difference, −0.024 mm). Of note, the beneficial effect of pioglitazone on mean CIMT was similar across pre-specified subgroups based on age, sex, systolic blood pressure, duration of DM, BMI, HbA1c value, and statin use.

Similar vascular beneficial effects of pioglitazone were demonstrated in the PERISCOPE study in which 543 patients with T2DM and CVD underwent coronary intravascular ultrasonography and were randomized to receive glimepiride, 1–4 mg, or pioglitazone, 15–45 mg, for 18 months, with titration to maximum dosage if tolerated [54]. Atherosclerosis progression was measured by repeat intravascular ultrasonography in 360 patients. The main outcome measure of change in percent atheroma volume (PAV) from baseline to study completion increased 0.73% with glimepiride and decreased 0.16% with pioglitazone (\(p = 0.002\)). Pioglitazone use was also associated with a slightly greater reduction in HbA1c (−0.55% versus 0.36% with glimepiride). In the pioglitazone group, HDL-C increased more (5.7 mg dL\(^{-1}\) vs. 0.9 mg dL\(^{-1}\)); and median triglyceride levels were lower by 16.3 mg dL\(^{-1}\) versus an increase of 3.3 mg dL\(^{-1}\) with glimepiride. Hypoglycemia was more common in the glimepiride group, and edema, fractures and decreased hemoglobin levels occurred more frequently in the pioglitazone group. In this study, in patients with T2DM and CAD, treatment with pioglitazone resulted in a significantly lower rate of progression of coronary atherosclerosis compared with glimepiride.

Given the beneficial effects of pioglitazone on surrogate measures of CVD, it is expected that these effects would result in lower cardiovascular morbidity/mortality. This was studied in the PROActive Study in which 5238 patients with T2DM and evidence of pre-existing CVD were randomized to pioglitazone 45 mg daily, or placebo, in addition to their usual glucose-lowering medications [47]. In order to assess the effect of pioglitazone on CAD, independent of its glucose-lowering effects, all patients were treated to optimal glucose, lipid and blood pressure goals. After an average follow up of ~3 years, pioglitazone treatment was associated with a modest 10% reduction in the risk of the primary composite endpoint of all-cause mortality, nonfatal myocardial infarction, stroke, acute coronary syndrome, and revascularization or amputation (\(p = 0.09\)). However, the “main secondary endpoint” of all-cause mortality, myocardial infarction, and stroke (defined before the unblinding of the data) was significantly reduced by 16% (\(p = 0.027\)). Of note, pioglitazone treatment was associated with an increase in congestive heart failure (CHF) and hospitalization for CHF. However, the criteria for heart failure was not clearly defined and it is unclear whether the frequency of this diagnosis was confounded by an increased presence of peripheral edema in the pioglitazone group. Of note, mortality from heart failure was not increased.

Although the PROActive study results demonstrated a reduction in all-cause mortality, myocardial infarction and stroke with pioglitazone, the study was not designed to determine the mechanism(s) for this benefit. This could have resulted from the lower HbA1c of 0.5% in the pioglitazone group, a small but significant reduction in blood pressure, or changes in the lipid profile. Given the multiple beneficial effects of pioglitazone in improving insulin sensitivity and traditional/nontraditional risk factors, it is tempting to speculate that these anti-atherogenic effects contributed to the results.

Other potential beneficial nonglycemic effects

Nonalcoholic steato-hepatitis (NASH)

Notwithstanding the fact that troglitazone was withdrawn due to liver toxicity, it does not appear that hepatotoxicity is a class effect with PPAR\(\gamma\) agonists. On the contrary, the majority of data suggests overall improvement in liver enzymes with pioglitazone use and possible beneficial effects in patients with NASH in whom insulin resistance is a characteristic feature. In a 6-month pilot study in 55 patients with IGT/T2DM and biopsy-confirmed NASH, pioglitazone 45 mg d\(^{-1}\) in conjunction with a hypocaloric diet improved glycemic control and also decreased liver fat content, hepatocyte ballooning/inflammation, and normalized liver aminotransferase levels [55]. In the larger PIVENS study in patients without diabetes, pioglitazone, however, was not superior to placebo for the composite histologic primary endpoint of standardized scores for steatosis, lobular inflammation, hepatocellular ballooning, and fibrosis. However, a post-hoc analysis of patients with well-defined NASH revealed that pioglitazone significantly improved steatosis and inflammation, compared to placebo [56].

Cancer

It is well known that compared with those without diabetes, patients with T2DM are at increased risk of several types of cancer, including a 40% increased risk of bladder cancer [57]. This strong association with cancer is hypothesized to be driven in part by shared risk factors (e.g. age, obesity, smoking) or perhaps by hyperinsulinemia, whereby elevated insulin levels in T2DM stimulate insulin receptors on neoplastic cells and promote cancer growth and division. There is strong evidence that the TZDs through their action on PPAR\(\gamma\) receptors in cancer cells suppress the growth of several human cancer lines, \textit{in vitro} and \textit{in vivo} [58], possibly by inhibiting primary tumor growth and metastasis by both direct and indirect anti-angiogenic effects [58]. Recently, however, there is concern that pioglitazone use may be associated with an increased
risk for bladder cancer and the drug has been withdrawn from clinical use in France and Germany. Pioglitazone is still available in the US and other parts of the world, with the warning about the association with bladder cancer [58]. As mentioned previously, interim data from the 8-year follow-up of a Kaiser Permanente study in the US of 193,099 eligible men and women with diabetes mellitus suggest that in the cohort analysis, there is no overall statistically significant increased risk of bladder cancer among patients ever treated with pioglitazone (http://general.takedapharm.com/Trial-Disclosure/01-03-TL-OPI-524-8-year-Interim-Report.pdf).

**Polycystic ovary syndrome (PCOS)**

PCOS is the most common cause of hyperandrogenemia in women and insulin resistance is a prominent feature of this syndrome. In clinical practice, metformin has been used as an insulin sensitizer in patients with PCOS. However, several studies have documented that the TZDs also have insulin-sensitizing and ovulation-inducing effects in women with PCOS [59]. In a recent meta-analysis of 161 trials, it was found that compared to metformin, pioglitazone was significantly more effective at reducing fasting insulin levels and improving the HOMA-IR index. However, pioglitazone was significantly less effective than metformin at reducing BMI. The effect of pioglitazone on fasting glucose levels, testosterone levels, and clinical measures of hyperandrogenism was no different from that of metformin [59]. Of note, although metformin and pioglitazone improve metabolic and hyperandrogenic parameters, their effects on improving reproductive outcomes is not clear. Given the possibility that pioglitazone may result in ovulation in some premenopausal anovulatory women, caution should be used with pioglitazone in PCOS women.

**Inflammation and immunity**

In addition to being present in the classic insulin-responsive tissues, PPARγ is also found in other cell types including the white and red pulp of the spleen, human bone-marrow precursors, macrophages/monocytes and colonic epithelium [60]. In vivo and in vitro studies have shown that the TZDs exert a largely anti-inflammatory effect through several mechanisms including inhibition of proliferation of activated T cells, inhibition of IL-2 secretion and/or the induction of apoptosis, and inhibition of the NF-kappaB pathway [60]. However, one problem in the interpretation of many of the studies of PPARγ effects on inflammation and immunity is that ligands thought to be specific for PPARγ may have regulatory effects on inflammatory parameters that are PPARγ-independent. Recent data suggest that PPARγ is the major functional receptor mediating the common aminosalicylate activities in inflammatory bowel diseases. There is also considerable interest is the recent discovery that some commensal bacteria and natural ligands present in food may induce PPARγ expression and activation in the colon [61].

**Dose/drug interactions**

**Dosage**

The usual starting dose for pioglitazone is 15 mg or 30 mg PO once daily [62]. For patients who respond inadequately to the initial dose, the dose can be increased in 15 mg increments up to 45 mg PO once daily. Patients not responding adequately to monotherapy with pioglitazone should be considered for combination therapy with other antidiabetic agents. The safety and efficacy of pioglitazone in adolescents and children has not yet been established. Some authorities recommend limiting pioglitazone to submaximal doses unless there is clear evidence of enhanced efficacy with higher doses in an attempt to minimize the adverse effects of the drug. This approach has not been evaluated in long-term studies.

**Drug interactions**

Pioglitazone is metabolized by the cytochrome P450 isoform CYP3A4 [62]. Safety and efficacy could be affected when pioglitazone is co-administered with drugs metabolized by this enzyme and blood glucose should be monitored more carefully. Closer blood glucose monitoring is also recommended in patients receiving pioglitazone in combination with CYP3A4 inhibitors like ketoconazole or itraconazole. Strong CYP2C8 inhibitors (e.g. gemfibrozil) increase pioglitazone concentrations and in these patients, the dose of pioglitazone should not exceed 15 mg daily [62]. On the other hand, CYP2C8 inducers (e.g. rifampin) may decrease pioglitazone concentrations. If an inducer of CYP2C8 is used during treatment with pioglitazone, changes in diabetes treatment may be needed based on clinical response without exceeding the maximum recommended daily dose of 45 mg.

Since pioglitazone is metabolized by CYP3A4, it is possible that pioglitazone may reduce the bioavailability of oral contraceptives containing ethinyl estradiol and norethindrone by induction of CYP3A4 and reduce the effectiveness of these agents. Caution should be exercised in these patients and alternative methods of contraception should be recommended.

**Monitoring requirements/precautions/contraindications**

**Renal impairment**

Renal elimination of pioglitazone is negligible, and the drug is excreted into the bile either unchanged or as metabolites and eliminated in the feces. No dose adjustment in patients with renal impairment is required [62].

**Hepatic impairment**

Patients with T2DM often have ALT elevations due to fatty liver disease or cardiac disease with episodic congestive heart failure. Hence, LFTs should be obtained before starting pioglitazone. In patients with abnormal liver tests, pioglitazone should be
initiated with caution. In patients with serum ALT greater than three times the reference range with serum total bilirubin greater than two times the reference range without alternative etiologies, pioglitazone should not be started since these patients are at risk for severe drug-induced liver injury. In addition, LFTs should be measured whenever clinically indicated [62].

**Congestive heart failure (CHF)**

Although most patients tolerate TZD treatment well, some patients will develop clinically significant edema. Experience from clinical trials indicate that there is a small expansion in plasma volume associated with TZD use which might result in edema occurring in the peripheral interstitial space [62]. In susceptible patients, the edema starts early, usually within several weeks, and is mild. Occasionally, it can be severe with rapid weight gain. Although the fluid accumulation usually occurs in the lower extremities, it may be seen throughout the body and rarely even precipitate pulmonary edema [62].

Based on the above, pioglitazone should be used with caution in patients with pre-existing edema, especially in those who have milder degrees of heart failure (NYHA Class 1 and 2) [62]. In patients with NYHA Class 3 and 4 cardiac status, treatment with pioglitazone is not recommended. In all other patients, pioglitazone therapy should be initiated with a low dose and patients should be evaluated early for edema (within several weeks). If edema does occur, dose reduction may be attempted. In those who develop symptoms of CHF, however, it may be prudent to discontinue the drug altogether, since in the published case reports of pulmonary edema, clinical improvement did not occur with diuretics and inotropes, while the TZD agent was still continued. In those with peripheral edema who do not respond to conventional doses of diuretics, it may be wise to discontinue the TZD agent permanently and use other oral agents or insulin. There is evidence that diuretics acting in the distal collecting ducts such as hydrochlorothiazide and spironolactone may mitigate the fluid retention associated with glitazone therapy, whereas proximally acting furosemide does not [63].

**Side effects**

Pioglitazone has now been in clinical use worldwide with several million patients treated for several years. The vast majority of these patients have tolerated these agents well and have shown clinical improvement in their glycemic status. During double-blind clinical trials, the most frequently reported adverse reactions in pioglitazone-treated patients were headache, hyperglycemia, myalgia, pharyngitis, sinusitis, tooth disorder, and upper respiratory tract infection [62]. Overall, the types of adverse experiences reported when pioglitazone was used either as monotherapy or in combination with sulfonylureas, metformin, and insulin were similar to placebo, except for hypoglycemia and edema. Importantly, the incidence of withdrawals from clinical trials due to an adverse event other than hypoglycemia was similar for patients receiving placebo or pioglitazone.

**Edema**

In clinical trials and postmarketing surveillance, the incidence of edema has varied from about 3% to 7.5% with the TZDs compared to 1% to 2.2% with placebo or other oral antidiabetic agents [64]. Of note, the highest incidence of edema has been reported when pioglitazone is used in combination with insulin. In controlled clinical studies, edema occurs in 15.3% in those treated with insulin+pioglitazone compared to 7.0% in insulin+placebo-treated patients. In a small minority of patients, the TZDs lead to significant peripheral edema and in some patients a precipitation/worsening of CHF [64].

It is still not clear by what precise mechanism(s) the TZDs cause edema or whether the edema is related to decompenasa- tion of cardiac function [64]. In early studies in nondiabetic volunteers, the TZDs were found to increase plasma volume by about 6–7% [64]. Whether it is this increase in plasma volume that leads to edema or other causes such as increased renal tubular sodium reabsorption, reflex sympathetic activation, alteration of intestinal ion transport, or increased production of VEGF (a potent tissue permeability factor) is still not clear (Figure 44.4). In rodent models, renal collecting duct-specific deletion of the PPARγ receptor has been shown to block TZD-induced fluid retention. In a recent study in obese patients with T2DM and poor glycemic control, it was noted that intensification of insulin therapy alone and the addition of pioglitazone both significantly improved glycemia (HbA1c 7.8 to 7.2% and 7.6 to 7.1%) and increased body weight (1.7 and 4.9 kg, respectively) [65]. Despite the greater weight gain with the addition of pioglitazone, the increase in body water was similar in both groups 1.6 versus 1.7 L. However, the increase in total body water was mainly extracellular and interstitial with pioglitazone compared with intracellular increase with insulin therapy alone. Pioglitazone also increased the filtered

---

**Figure 44.4** Multifactorial mechanisms of edema formation with thiazolidinediones.
load of sodium reabsorbed at the distal nephron with no net change in the fractional excretion of sodium. The authors speculated that since TZD treatment (possibly through a PPARγ effect) may cause increased collecting duct reabsorption of sodium, this potential edema-causing effect may be offset by decreased proximal-tubular reabsorption of sodium. Those diabetics who do not have this potential renal compensatory mechanism may be more prone to develop fluid accumulation and heart failure. As mentioned previously, fluid retention may be minimized by co-administration of thiazide diuretic or spironolactone. The recent development of partial PPARγ agonists with potent antidiabetic activity, but no fluid retention, weight gain or bone loss properties is a promising development in this field [11].

**Weight gain**

In clinical studies, pioglitazone treatment has been accompanied by weight gain in a dose-related manner [62]. The change in average weight in US placebo-controlled monotherapy trials ranged from 0.5 kg to 2.8 kg for pioglitazone-treated patients compared to a weight loss of 1.3 kg to 1.9 kg in placebo-treated patients. In combination with a sulfonylurea, the change in average weight was 1.9 kg and 2.9 kg for 15 mg and 30 mg of pioglitazone, respectively, and −0.8 kg for placebo. In combination with metformin, the increase in average weight was 1.0 kg for 30 mg of pioglitazone versus −1.4 kg for placebo. The change in average weight in combination with insulin was 2.3 kg and 3.7 kg for 15 mg and 30 mg of pioglitazone, respectively, and no weight change for placebo. It is also important to note that thiazolidinedione-induced weight gain is due to a combination of fluid retention and adipose tissue accumulation. As discussed earlier [65], the increase in total body water was mainly extracellular and interstitial with pioglitazone compared with intracellular increase with insulin therapy alone. Also, with pioglitazone, there is an increase in the subcutaneous adipose tissue compartment and lean body mass along with a trend towards a decrease in visceral adipose tissue [41].

**Bone fractures**

Data from in vitro and rodent studies demonstrate that PPARγ activation inhibits bone formation by diverting mesenchymal stem cells from the osteogenic to the adipocytic lineage and also possibly by increasing bone resorption secondary to stimulating osteoclast development. PPARγ agonists may also have indirect adverse skeletal effects through modulation of circulating hormones/cytokines known to influence bone metabolism [66]. In the PROactive study [47], the incidence of bone fracture in females was 5.1% for pioglitazone versus 2.5% for placebo. This difference was noted after the first year of treatment and persisted during the course of the study. The majority of fractures in female patients were nonvertebral (lower limb and distal upper limb). No increase in the incidence of fracture was observed in men. The risk of fracture should be considered in the care of patients, especially female patients, treated with pioglitazone and attention should be given to assessing/maintaining bone health per current standards of care.

**Anemia**

It is well documented that the TZDs increase plasma volume and due to a dilutional effect, there will be decreases in hemoglobin. This has been shown to occur in a dose-related fashion in patients treated with pioglitazone alone or in combination [62]. These changes occur primarily during the first 4–12 weeks of therapy and remain relatively constant thereafter and are not associated with any significant hematologic clinical effects. In US double-blind studies, anemia was reported in 0.3 to 1.6% of pioglitazone-treated patients and 0 to 1.6% of placebo-treated patients in monotherapy and combination studies [62].

**Hypoglycemia**

The TZDs do not stimulate insulin secretion and hence when used as monotherapy, are not expected to cause hypoglycemia. However, hypoglycemia has been reported during combination therapy with sulfonylureas or insulin [62]. Hypoglycemia was reported for 1% of placebo-treated patients and 2% of patients receiving pioglitazone in combination with a sulfonylurea. In combination with insulin, hypoglycemia was reported for 5% of placebo-treated patients, 8% for patients treated with pioglitazone 15 mg, and 15% for patients treated with pioglitazone 30 mg.

During required laboratory testing in the pioglitazone clinical trials, sporadic, transient elevations in creatine phosphokinase levels without symptoms were observed [62]. These elevations resolved without any apparent clinical sequelae and the relationship of these events to pioglitazone therapy is unknown.

**Future developments**

Given the unfavorable side-effect profile of the TZDs, there has been considerable interest in developing PPAR agonists with glucose-lowering/insulin-sensitizing PPARγ agonist properties of the TZDs and the lipid-modifying PPARγ properties of fibrates along with an improved safety profile.

**Conclusion**

The introduction of the PPARγ agonist thiazolidinediones heralded a new era in the treatment of patients with T2DM. These agents exert direct effects on the mechanisms of insulin resistance which is a major pathophysiologic abnormality in T2DM. As highly selective and potent PPARγ agonists, the TZDs regulate the expression of numerous genes that affect carbohydrate/lipid metabolism and vascular function. This results not only in lower blood glucose levels and improved glycemia, but also amelioration of several components of
the insulin-resistance syndrome including dyslipidemia, hypertension, and endothelial dysfunction which untreated lead to accelerated atherosclerosis and premature cardiovascular morbidity/mortality in patients with T2DM. However, these beneficial effects have not translated into clear-cut cardiovascular benefits. Today, pioglitazone is the only agent in this class that is widely available in the US and most of the world. Pioglitazone treatment improves insulin sensitivity and glycemic/lipid parameters and also has favorable effects on diabetes prevention. In one study, pioglitazone reduced the composite secondary outcome of all-cause mortality, nonfatal myocardial infarction, and stroke in patients with T2DM and high CVD risk. However, pioglitazone use has been limited by concerns about weight gain, bone loss, fluid retention, edema, and a recent association with increased bladder cancer risk. Despite this, the PPARγ agonists still have a role in the treatment of T2DM, since this is the only class of drugs that primarily targets insulin resistance. Also, these drugs can be used in patients with renal impairment and exhibit a low incidence of hypoglycemia. In the future, it is possible to envisage the availability of tailored PPAR agonists which through their selective effects will have enhanced beneficial effects on glucose, lipid and vascular endothelial metabolism without the unwanted side effects of weight gain and fluid retention.

Acknowledgments

This work was supported by the Department of Veterans Affairs and the VA San Diego Healthcare System, San Diego, CA.

References


62 Actos (Prescribing information). Lincolnshire, IL: Takeda; August 2012.


CHAPTER 45

α-Glucosidase inhibitors

Josée Leroux-Stewart1, Rémi Rabasa-Lhoret1,2 and Jean-Louis Chiasson1,3

1Division of Endocrinology, Centre hospitalier de l’Université de Montréal (CHUM), Montréal, QC, Canada
2Department of Nutrition, Université de Montréal, IRCM, Montréal, QC, Canada
3Research Centre, Centre Hospitalier de l’Université de Montréal (CRCHUM), Montréal, Canada

Key points

• Alpha-glucosidase inhibitors reduce postprandial plasma glucose rise by delaying glucose absorption by the small intestine.
• Alpha-glucosidase inhibitors are effective in improving HbA1c whether they are prescribed as monotherapy or in combination with other antidiabetic medications.
• The major side effects of alpha-glucosidase inhibitors are gastrointestinal in the form of flatulence, abdominal cramps and rarely diarrhea; these side effects can be minimized by the “start low go slow” policy.
• Acarbose and voglibose have been shown to be effective in reducing the risk of diabetes in subjects with IGT.
• Studies have suggested that acarbose could reduce the risk of cardiovascular disease in subjects with IGT and type 2 diabetes mellitus.
• Alpha-glucosidase inhibitors have an excellent safety profile.

Mechanism of action

α-Glucosidase inhibitors competitively block small intestine brush border α-glucosidases that are necessary to hydrolyze di-, oligo-, and polysaccharides to monosaccharides for absorption [4]. Normally, carbohydrates are primarily and rapidly absorbed in the small intestine. With AGI, carbohydrate absorption and digestion occur throughout the small intestine, resulting in a slower absorption and a blunting of the postprandial rise in plasma glucose [5].

Three AGI have been developed (acarbose, miglitol, and voglibose) with similar pharmacologic profiles (Figure 45.1) [6,7]. Acarbose, a pseudotetrasaccharide of microbial origin, is structurally analogous to an oligosaccharide derived from starch digestion. It is made up of a maltose unit linked to an acarviosine unit, which represents the active part of the molecule [5]. The acarviosine unit has nitrogen linkage, which is responsible for its high affinity for the carbohydrate-binding site of various α-glucosidases, and exceeds the affinity (10- to 100,000-fold) of regular oligosaccharides from nutritional carbohydrates [4]. Because of this C-N linkage, acarbose cannot be cleaved and, therefore, enzymatic hydrolysis is blocked [4]. Despite its high affinity for these enzymes, acarbose binding is reversible, and its inhibition kinetic is competitive. Owing to patients, however, find these regimens difficult to follow. The α-glucosidase inhibitors (AGI) specifically delay the digestion of complex carbohydrates and decrease the postprandial rise in plasma glucose, thus reproducing the effect of a low glycemic index/high-fiber diet. These drugs significantly reduce postprandial glycemic and insulineic excursions whether they are used as monotherapy or in combination for the treatment of T2DM and also in T1DM. These drugs can be associated with gastrointestinal symptoms but have an excellent safety profile.

Introduction

Dietary and lifestyle interventions still remain the cornerstone in the treatment of patients with type 2 diabetes mellitus (T2DM). However, when diet and physical exercise fail to achieve optimal glycemic control, oral hypoglycemic agents must be added. Most of these agents have been shown to decrease fasting plasma blood glucose, but postprandial hyperglycemia persists in more than 60% of patients, accounting in part for the sustained increase in HbA1c [1]. It has been suggested that postprandial glycemic excursions contribute to the development of diabetes-specific complications (i.e., retinopathy and nephropathy) and, based on epidemiologic studies, could also be involved in the development of macrovascular complications [2,3].

Slowly absorbable carbohydrates and high-fiber diets blunt the postprandial increase in plasma glucose and insulin levels. Most
its specificity for α-glucosidases, β-glucosidases (e.g. lactases) are not inhibited by acarbose, and therefore lactose absorption is not affected. Intestinal glucose absorption is also not affected [4]. Because less than 1–2% of the active compound is absorbed, acarbose displays its inhibitory activity throughout the small intestine up to the ileum [4]. Acarbose is cleaved in the large intestine by bacterial enzymes into several metabolizable intermediates and 4-methylpyrogallol derivatives. Thus even though acarbose itself is poorly absorbed, 35% of an oral dose appears as metabolites in the urine [8].

Chemically, miglitol closely resembles the glucose molecule and is significantly absorbed through a jejunal transport mechanism identical to that for glucose. It then circulates and concentrates in enterocytes of the small intestine where it competes for the α-glucosidases. It is not metabolized and is excreted quantitatively by the kidney [8].

Voglibose is an N-substituted derivative of valiolamine, a branched-chain aminocyclitol isolated from the fermentation broth of Streptomyces hygroscopicus, subspecies limoneus. Its N-substituted moiety is derived from glycerol. Voglibose is poorly absorbed by the gastrointestinal tract and hardly metabolized [9].

To be efficient, AGI must be present at the site of enzymatic action at the same time as the polysaccharides, oligosaccharides, or disaccharides. Therefore, they should be taken with the first bite and not more than 15 min after the beginning of the meal [10].

**Clinical efficacy of α-glucosidase inhibitors**

**Treatment of diabetes mellitus**

Much experience has been gained with AGI in the treatment of diabetes mellitus. The benefit of adding them to the diet alone or in combination with other oral agents, as well as with insulin in type 2 or type 1 diabetic patients has been investigated in numerous randomized, double-blind, placebo-controlled trials, as well as open-labeled controlled studies, and postmarketing surveillance trials.

In a Cochrane systematic review of 22 placebo-controlled studies (n = 2238), acarbose showed a significant reduction in postprandial plasma glucose (PPG) of −2.32 mmol L⁻¹ (95% CI −2.73, −1.92) as well as in fasting blood glucose (FBG) of 1.09 mmol L⁻¹ (95% CI −1.36, −0.83) (28 studies, n = 2838) [11]. Through their effects on FBG and PPG, AGI significantly impact on HbA1c. Multiple studies have shown that the decline usually ranges from 0.5 to 1.5% (Figure 45.2). In fact, the results of the same Cochrane meta-analysis of 28 studies comparing the effect of acarbose versus placebo showed a decline in HbA1c of 0.77% (95% CI −0.9%, −0.64%). The total effect of miglitol versus placebo in seven studies was a decline of 0.68% (95% CI −0.93%, −0.44%) [12]. Few large-scale studies have been done with voglibose, but when added to current treatment, results show a mean HbA1c reduction of 0.31% [13–18].

Large, controlled or open trials have revealed a number of constant features associated with acarbose therapy in
α-Glucosidase inhibitors

Diet

1.5
1.0
0.5
0
–0.5
–1.0
–1.5
Metformin

1.5
1.0
0.5
0
–0.5
–1.0
–1.5
Sulfonylurea

1.5
1.0
0.5
0
–0.5
–1.0
–1.5
Insulin

0
3
6
9
12
Time (months)

Figure 45.2 Effects of acarbose and placebo on HbA1c levels throughout the 1-year study period in groups of patients treated with diet alone, metformin, sulfonylurea, or insulin. Data are expressed as means with 95% CI. *

Diabetes [19–21]. The glucose-lowering effect is observed from the first week of therapy, and is maintained throughout the treatment period, even after 5 years. Acarbose efficacy is present whatever the current treatment. However, a greater response to treatment has been reported in patients with recently diagnosed diabetes and/or naïve to oral hypoglycemic agents. The drug is effective independent of sex or ethnicity. No weight gain occurs and most studies have shown small but consistent weight loss. Used as monotherapy, there is no risk of hypoglycemia. The main cause of treatment discontinuation is gastrointestinal side effects, with discontinuation rates varying from 2% to 40%. However, very few serious adverse effects are associated with AGI, making it a particularly safe drug.

Dose–response relationship

The dose–response relationship has been shown, mostly with acarbose, in large multicenter trials at dosages ranging from 25 to 300 mg three times daily. Overall, these trials indicate that the 100 mg three times daily dose evokes a near maximal response [22,23].

Efficacy of α-glucosidase inhibitors combined with diet therapy in patients with type 2 diabetes

Major randomized controlled trials, up to 5 years’ duration, including type 2 diabetic patients insufficiently controlled by diet alone, have tested AGI versus placebo [11,19,20,22–25]. Overall, these studies confirmed the efficacy of AGI, showing a mean reduction of 0.66% in HbA1c, a 1.03 mmol L−1 decline in FBG, and a 2.6 mmol L−1 fall in PPG.

A small study with patients newly diagnosed with T2DM compared the effects of dietary treatment alone or with the addition of voglibose on abdominal visceral adipose tissue area (VAT) and subcutaneous adipose tissue area (SAT). Voglibose showed better results for both parameters after only 3 months of treatment. In the diet group, changes in VAT were of 27.7 ± 20.1 cm2 compared to 39.4 ± 20.4 cm2 in the voglibose group. For the SAT, changes of 20.9 ± 21.8 cm2 were observed for the diet group versus 29.3 ± 32.8 cm2 for voglibose [17].

Efficacy of α-glucosidase inhibitors in comparison to sulfonylureas or metformin

Ten studies have compared AGI head to head with sulfonylureas in patients with T2DM inadequately controlled on diet alone [26–34]. AGI appeared to be less potent, with a mean adjusted HbA1c reduction of 0.74% versus 1.18% with sulfonylureas.

Few studies have compared AGI with metformin in patients failing dietary treatment. In a study by Hoffman and Spengler, acarbose (300 mg d−1) had the same efficacy as metformin (1700 mg d−1), with a mean HbA1c reduction of 1%. Chiasson et al. showed that miglitol (300 mg d−1) had a lower efficacy (HbA1c -0.4%) than metformin (1500 mg d−1; HbA1c -1.2%) [24], but there was a synergism effect when combining miglitol with metformin (HbA1c -1.8%).

Efficacy of α-glucosidase inhibitors as an adjunct treatment to sulfonylurea and/or metformin

A number of studies have tested acarbose and miglitol as add on therapy in type 2 diabetic patients inadequately controlled with sulfonylureas [19,20,26,30–39]. When added to sulfonylurea therapy, HbA1c was decreased by an adjusted mean of 0.81%, FBG by 1.1 mmol L−1, and PPG by 2.25 mmol L−1. In patients previously treated with sulfonylureas, the benefit of acarbose
seemed comparable to the addition of metformin for lowering HbA1c [36,39].

In type 2 diabetic patients suboptimally controlled with metformin, the addition of acarbose produced a mean reduction of 0.85% in HbA1c, 1.36 mmol L⁻¹ in FBG, and 2.76 mmol L⁻¹ in PPG [19,20,40–43].

When combined sulfonylurea and biguanide therapy failed to achieve adequate metabolic control, the addition of acarbose resulted in a similar reduction in mean HbA1c of 0.81% [19,44–48]. In a large open trial, diabetic patients failing with a combination of sulfonylurea and biguanide had a much smaller HbA1c benefit than did patients treated previously with a single oral hypoglycemic agent [21].

**Thiazolidinedione and α-glucosidase inhibitors**

Few data are available for this combination. In one study, 274 patients on sulfonylurea and metformin were randomized to either pioglitazone or acarbose. Patients were followed for 9 months. The reduction in HbA1c was greater in the group treated with pioglitazone (−1.4 vs. −0.9%) but weight gain was also significantly higher (+1.4 kg vs. −1.2 kg) [45]. Likewise, a study of 60 patients randomized to either pioglitazone or voglibose for 6 months showed a slightly better performance on HbA1c in the pioglitazone group (−0.3 vs. −0.2%). However, no change in BMI or waist circumference was observed with pioglitazone, whereas these parameters significantly decreased with voglibose (−1.2 kg m⁻² in BMI and −3 cm in waist circumference) [49].

**Efficacy of α-glucosidase inhibitors in comparison to dipeptidyl peptidase-4 inhibitors**

The dipeptidyl peptidase-4 inhibitors (DPP-4 inhibitors) alogliptin, vildagliptin, sitagliptin, and linagliptin have been compared to voglibose or acarbose in a few double-blind randomized trials (n > 2200). Overall, the mean adjusted HbA1c reduction was of 0.54% in the AGI groups versus 0.9% in the DPP-4 inhibitor groups [14–16,50]. Both therapies are associated with a low hypoglycemic risk.

**Efficacy of α-glucosidase inhibitors in insulin-treated type 2 diabetic patients**

The addition of bedtime insulin is a frequent option for patients with T2DM insufficiently controlled with diet, exercise, and maximum tolerated doses of two or three oral hypoglycemic agents. The addition of acarbose has been compared to the introduction of bedtime NPH. Although acarbose resulted in improvement of metabolic control compared to placebo (HbA1c −0.8% vs. −0.2%), it was less effective than insulin (HbA1c −2.3%) [51]. Therefore, this strategy is generally considered as a compromise for patients who refuse insulin injections.

When acarbose was combined with insulin therapy, however, it resulted in significant metabolic improvement, with a mean adjusted reduction in HbA1c of 0.5% without any effect on insulin dosages but also without any weight gain [20,23,52–55].

**Comparative studies between α-glucosidase inhibitors**

Few studies have compared the different AGI amongst themselves. In nondiabetic subjects, voglibose seemed slightly less potent than acarbose, with no difference in side effects [7]. Miglitol gave softer stool and slightly less flatulence than acarbose [56]. In type 2 diabetic patients, voglibose was associated with less gastrointestinal side effects than acarbose, but was also less effective in reducing PPG excursion [6,57].

**Benefits of α-glucosidase inhibitors in elderly patients**

Because of its mechanism of action and safety profile, acarbose could be an interesting first-line drug for elderly type 2 diabetic patients because of the low hypoglycemic and drug interaction risks [5,25,30,58]. Furthermore, postprandial hypotension is an important clinical problem and has been recognized as a common cause of syncope and falls in the elderly population [59,60]. Interestingly accumulating data support the usefulness of AGI in attenuating this phenomenon. In fact, many studies and case reports have shown that these drugs reduce, sometimes markedly, the postprandial hypotensive response to carbohydrate intake in nondiabetic patients [61–63] and in type 2 or type 1 diabetic patients suffering from this problem [64].

Many mechanisms have been proposed, one of which is the effect of AGI on carbohydrate absorption. In fact, by slowing the rate of absorption, the acute rise in splanchnic blood flow is blunted [63]. Furthermore, constipation is often a major problem in this population, and so gastrointestinal side effects reducing colonic transit time may in some circumstances be perceived here as an advantage [30,65].

**Impaired glucose tolerance and prevention of type 2 diabetes**

The STOP-NIDDM trial has investigated the potential of acarbose in preventing or delaying the development of T2DM in subjects with impaired glucose tolerance (IGT). A total of 1429 subjects were randomized to acarbose or placebo. On the basis of a single OGTT (oral glucose tolerance test), acarbose reduced the risk of developing T2DM by 25% [66] (Figure 45.3). If two positive OGTTs were used to confirm the diagnosis of T2DM as recommended by the WHO (World Health Organization) and the Expert Committee of the ADA (American Diabetes Association), the relative risk reduction was 36.4%. The beneficial effect of acarbose was independent of age, sex, and BMI. The STOP-NIDDM trial also showed that acarbose treatment was additionally associated with a significant increase in the reversion of IGT to normal glucose tolerance. These results indicate that 11 patients should be treated for 3.3 years to prevent or delay one case of T2DM. However, those who had the metabolic syndrome (MS) (61%) had a higher risk of progressing to diabetes, but were also more responsive to acarbose; in fact, only
5.8 patients with prediabetes and the MS had to be treated to prevent one case of diabetes, an efficacy similar to that of lifestyle modification [67]. Acarbose treatment was also associated with a significant reduction in the incidence of macrovascular events and new cases of hypertension, two prespecified secondary endpoints (Figure 45.4) [68]. Thus, in subjects with IGT, acarbose is effective in decreasing the risk of diabetes and hypertension, and potentially, cardiovascular complications [68,69].

Similarly, a multicenter randomized, double-blind trial was performed in 1780 patients with IGT, randomly assigned to receive either voglibose or placebo [70]. Because of the highly significant reduction in the incidence of diabetes in the voglibose group, the study was discontinued prematurely after a mean duration of 48.1 weeks. Voglibose treatment was associated with a 40.5% reduction in the risk of developing T2DM compared to placebo (HR 0.595; 95% CI 0.433–0.818; \( p = 0.0014 \)). Voglibose also increased the conversion of IGT to normal glucose tolerance (HR 1.539; 95% CI 1.357–1.746; \( p < 0.0001 \)). Therefore, as with acarbose, voglibose is effective in decreasing the risk of diabetes in subjects with IGT.

The effect of AGI on the prevention of T2DM is very similar to that of metformin (31%) [71]. Lifestyle intervention, however, was more effective in reducing the risk of progressing from IGT to diabetes (58%) [71,72]. On the other hand, the thiazolidinediones, rosiglitazone and pioglitazone, were more effective than lifestyle (60% and 72%, respectively), but because of their adverse effects, will very unlikely be recommended for the prevention of diabetes [73,74].

**Efficacy of \( \alpha \)-glucosidase inhibitors in type 1 diabetic patients**

\( \alpha \)-Glucosidase inhibitors have been tested in patients with T1DM. These trials have shown acute and chronic reductions of PPG. In long-term treatment, eight published trials including 471 patients showed a mean adjusted HbA1c decrease of 0.54%
when acarbose was added to regular human insulin [75–78]. In most studies, acarbose improved the plasma glucose profile with lowered postprandial plasma glucose excursion and delayed nadir values [79,80]. Studies have also shown that when AGI are added to regular insulin, there is a decrease in the risk of nocturnal and daytime hypoglycemia as well as during postprandial exercise [80,81]. Though acute studies have shown an insulin-sparing effect of acarbose (approximately 30%), especially for pre-meal insulin requirements [75,79,82], the largest placebo-controlled trial in patients with T1DM could not confirm this sparing effect after 24 weeks of treatment [78]. The usefulness and the interests of the addition of AGI in T1DM have disappeared since the rapid-acting insulin analogues have become available.

**Other effects of α-glucosidase inhibitors**

**Gastrointestinal peptide secretion**

α-Glucosidase inhibitors modify the secretion of gastrointestinal peptides, such as glucose-dependent insulino-tropic polypeptide (GIP) and glucagon-like peptide 1 (GLP-1). When the AGI are taken with a meal rich in carbohydrates, GIP secretion is decreased while GLP-1 is markedly increased, especially in the late postprandial period [8,82–85]. Whether these effects on GIP and GLP-1 play a role in the therapeutic effect of AGI is unclear.

**Cardiovascular risk factors**

In animal models of diabetes and/or dyslipidemia, AGI reduce serum triglycerides in a dose-dependent manner. In humans, acarbose was shown to lower the postprandial triglyceride rise [40,86,87] but its effect on fasting levels has been inconsistent with most studies reporting a small but significant reduction [28,39,40,75,86,88–93].

The effects of AGI on fasting total cholesterol and HDL cholesterol appear to be minor or marginally significant [18,21,88,89,93]. The reported changes may reflect modifications in glycemic control more than a direct action of the drug [8].

Because of their effect on PPG and insulin excursion, AGI have been shown to decrease insulin resistance in most [25,58,66,94–97] but not all studies [98–100]. The negative results in the latter studies may be related to the very poor glycemic control.

Acarbose has been shown to have a favorable effect on blood pressure. Hanefeld et al. performed a meta-analysis of seven randomized, double-blind, placebo-controlled acarbose trials, for a total of 2180 type 2 diabetic patients, measuring the effects of acarbose on metabolic parameters as well as cardiovascular events [101]. Treatment duration varied from 52 to 164 weeks. The overall analysis revealed that systolic blood pressure was significantly lowered by 2.7 mmHg (p = 0.024) with acarbose treatment. Furthermore, data from the STOP-NIDDM trial indicate that acarbose could delay the appearance of new cases of high blood pressure in IGT patients [68].

Acarbose has been shown to have a neutral or a lowering effect on body weight. Two different studies could not show any effect of acarbose on energy intake, nutrient intake, or dietary patterns [102,103]. Theoretically, carbohydrate malabsorption associated with acarbose could produce weight loss. However, carbohydrates that reach the colon are metabolized by bacteria into short chain fatty acids which are then absorbed, resulting in no or minimal caloric loss. In general, clinical trials with acarbose found no [19,104] or small (~1 kg) weight loss after 24 weeks to 5 years of treatment [6,21,28,36,40,44,50,66,102,105]. Overall, AGI exert neutral or slightly favorable actions on different cardiovascular risk factors that are part of the metabolic syndrome: excess body weight, high blood pressure, postprandial hyperlipemia, and insulin resistance.

**Potential effects of α-glucosidase inhibitors on the prevention of long-term diabetic complications**

Data from animal studies suggest that acarbose has significant beneficial actions on microvascular complications [106]. All studies indicate that the positive effects are through its influence on glycemic control. Studies in different animal models of diabetes have shown that acarbose successfully reduces or prevents renal abnormalities, eye problems, and neuropathy associated with diabetes [106].

Finally, there is rising evidence that acarbose could be useful in reducing macrovascular complications associated with diabetes [106]. In a rodent model (the LA-corpulent rat), 8 months of acarbose treatment led to a significant regression in myocardial lesions. Similar findings on aortic atherosclerosis have been obtained in a rabbit model fed an atherogenic diet. Carotid artery intima-media thickness (IMT) is generally accepted as a surrogate for atherosclerosis and is recognized as an independent predictor of coronary heart disease and stroke [107]. Since postprandial hyperglycemia has been associated with increased common carotid IMT [108,109], much interest has emerged regarding the potential effect of an AGI on this surrogate measurement. Five recent randomized controlled trials of at least 1 year duration studied the progression of carotid IMT with an α-glucosidase inhibitor in comparison to a placebo, no treatment or other hypoglycemic agents, in 411 patients with IGT or T2DM [18,88,89,104,110]. All five studies were associated with a significant decrease in the annual progression of carotid IMT when treated with one of the AGI, suggesting an anti-atherosclerotic effect in humans.

In the STOP-NIDDM trial, acarbose treatment in subjects with IGT was associated with a 49% reduction in cardiovascular events [68]. This was further supported by a meta-analysis of seven randomized, double-blind, placebo-controlled acarbose trials in patients with T2DM treated from 52 to 164 weeks (n = 2180) [101]. The analysis showed a highly significant relative risk reduction of 35% for any cardiovascular event in the acarbose-treated group.
However, it is still debated whether the reduction in cardiovascular events associated with AGI treatment in subjects with IGT or diabetes is a cause-and-effect relationship as suggested by the STOP-NIDDM trial and the MeRIA meta-analysis [3,68]. It is hoped that the question will be answered by the Acarbose Cardiovascular Evaluation (ACE) trial, a double-blind, randomized, multicenter, cardiovascular intervention study currently in progress. In this study, 7500 patients with IGT and coronary heart disease are randomized to either acarbose or placebo and followed for a minimum of 4 years. The primary objective is to determine whether reducing postprandial hyperglycemia with acarbose can reduce cardiovascular-related morbidity and mortality [111].

The exact mechanisms by which PPG could contribute to diabetic vascular complications are still being studied. Nonetheless, there is a growing literature suggesting that the generation of reactive oxygen species may be the common pathway [112,113] triggered by acute hyperglycemia resulting in oxygen stress, subclinical inflammation, activation of procoagulation factors and endothelial dysfunction, the first step to atherosclerosis. The combination of all these atherogenic mechanisms could potentially cause CVD [114].

**Treatment of reactive hypoglycemia**

On the basis of their pharmacology, AGI can be useful in the treatment of reactive hypoglycemia and/or dumping syndrome in nondiabetic patients [8,115,116].

**Tolerability of acarbose**

Thousands of patients have been treated with this medication for several years without any major adverse effects (Table 45.1); acarbose is a safe drug [8].

**Gastrointestinal side effects**

The main side effects of treatment are gastrointestinal symptoms consisting of abdominal distention, flatulence, diarrhea, and borborygmus in approximately 50% of patients. These adverse events are related to intracolonic fermentation and the consequent gas production of carbohydrates not absorbed in the small bowel. The symptoms are usually mild to moderate and tend to decrease with continued treatment [22,25]. With time, the α-glucosidases of the distal small intestine will be activated, and more carbohydrates will be digested and absorbed distally, reducing the spill over into the large bowel. That is why starting with a low dose and increasing it slowly can help minimize the side effects [8,66]. The importance of a “start low go slow” strategy is confirmed by the positive relationship between higher initial doses and higher discontinuation rates.

There is no correlation between carbohydrate intake, fiber content of the diet or meal pattern (regular vs. irregular) and side effects associated with acarbose [117].

The treatment discontinuation rate, close to 25% in most large trials, is mostly related to gastrointestinal side effects [20,24,66]. Some studies including the UKPDS have reported much higher discontinuation rates (~40%) [8,19]. However, others have found a much lower rate of patients unable to tolerate AGI which is comparable to that observed with biguanides (i.e., 5–7%) [22,24,36,118]. No clear explanations are available for such discrepancies, but false evaluation of the gastrointestinal side effects can lead to unnecessary gastroenterologist consultations.

Despite these gastrointestinal symptoms, the quality of life of patients on acarbose has been shown to be as good as, if not better, than with sulfonylurea or insulin therapy [25].

**Hypoglycemia**

Acarbose monotherapy does not cause hypoglycemia [19]. However, it may potentiate the hypoglycemic action of sulfonylureas and insulin. A reduction in dosage of concomitant hypoglycemic agents may be necessary when acarbose is introduced. However, if a patient taking acarbose experiences hypoglycemia, it is recommended that it should be treated with glucose since the absorption of sucrose and complex carbohydrates could be delayed by the drug.

---

**Table 45.1** Comparison of adverse effects of different available antidiabetic oral agents

<table>
<thead>
<tr>
<th>Adverse effect</th>
<th>Sulfonylureas</th>
<th>Metformin</th>
<th>Thiazoles</th>
<th>Acarbose</th>
<th>DPP-4 inhibitors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight gain</td>
<td>+</td>
<td>–</td>
<td>+/++</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Hypoglycemia</td>
<td>+/++</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Hypersensitivity reactions</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>Drug interactions</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>Lactic acidosis</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Gastrointestinal disturbance or hepatic reaction</td>
<td>+</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>–</td>
</tr>
<tr>
<td>Edema</td>
<td>–</td>
<td>–</td>
<td>++</td>
<td>–/+++</td>
<td>–</td>
</tr>
</tbody>
</table>

*Minor clinical problem
**Major clinical problem
*Not or rarely associated with this side-effect
Other potential and demonstrated side effects related to α-glucosidase inhibitor therapy

Pneumatosis cystoides intestinalis also known as pneumatosis coli, pseudolipomatosis or intestinal emphysema is a rare condition in which gas is abnormally found in a linear or cystic form in the submucosa or subserosa of the bowel wall. Clinically, this disorder has been associated to a large variety of conditions and so its relevance must be interpreted within the clinical context. It can be suggestive of a severe underlying condition such as bowel infarction or perforation but can also be a benign incidental finding. Interestingly, since 1999, all three AGI have been associated with the typical radiographic finding. Many mechanisms have been proposed to explain the causality behind this association. It is thought to mainly involve raised intraluminal pressure due to gas production through fermentation of nonabsorbed carbohydrates. Mucosal breaks could then occur, allowing gas to dissect through tissues of the bowel wall.

Thirty case reports of pneumatosis cystoides intestinalis in patients treated with AGI have been documented [119–122]. The onset occurred from 1 week to many years after starting treatment. Patients complained of abdominal pain or distension, and/or rectal bleeding. Four patients were asymptomatic and this was an incidental finding. Conservative treatment was employed in most cases, but exploratory surgery was judged appropriate for seven patients. Treatment consists of immediate discontinuation of the drug with, in some cases, the use of antibiotics or the administration of oxygen. All thirty cases reported complete clinical and radiologic healing usually within the month following the discontinuation of the drug.

Because of gastrointestinal side effects, there has been concern about the risk of colon cancer associated with AGI. There has been no increase of colon cancer in large cohorts followed for more than 1 year [19,20,66]. Furthermore, in a large cohort study of 39515 newly diagnosed type 2 diabetic patients of which 2918 were treated with AGI and followed for seven years, AGI treatment was associated with a lower risk of hepatic cancer (adjusted HR, 0.62; 95% CI, 0.4-0.94) [123].

A few case reports have described the association of ileus with acarbose treatment in Japanese patients [124].

In approximately 3 million patient-years of international postmarketing experience with acarbose, 62 cases of serum transaminase elevations >500 IU L⁻¹ have been reported [125]. Of these 62 patients, 41 were treated with doses of 100 mg three times daily or greater. Interestingly, 33 of 45 patients for whom weight was reported weighed less than 60 kg. Therefore, the maximum recommended dose for patients <60 kg is 50 mg three times daily, increasing to 100 mg three times daily maximum for patients >60 kg. For 55 of the 59 cases where follow-up was recorded, hepatic abnormalities improved or resolved upon discontinuation. Moreover, two cases of severe hepatoxic reaction categorized as probable idiosyncratic reaction related to acarbose have been reported [126,127]. Thus, hepatic reactions are rare and unpredictable.

The occasional observation of anemia is controversial and probably not of clinical significance. Finally, occasional reports indicate that acarbose could lower or increase digoxin plasma levels, and so monitoring should probably be done when AGI therapy is initiated or the dose is modified [128].

Overall, gastrointestinal side effects are the main clinical issue. To maximize treatment compliance, it is essential for clinicians to start patients on a low dose, titrating the dose gradually based on efficacy and tolerance, and make sure to discuss these anticipated side effects with the patients as well as their expected decrease over time.

Contraindications to α-glucosidase inhibitors

Formal contraindications to acarbose therapy are intestinal malabsorption syndromes, inflammatory bowel disease, colonic ulceration, partial intestinal obstruction, and cirrhosis. It is also contraindicated in cases of severe renal impairment (creatinine clearance <25 mL min⁻¹), in pregnant or lactating women, and in children below 12 years of age, because of a lack of data in these groups. Interestingly, however, reproduction studies have been performed in rabbits at doses up to 32 times those given to humans and have revealed no evidence of teratogenicity [125].

Since trials in diabetic cirrhotic patients have demonstrated a positive action of acarbose on glycemic control without adverse effects on liver function, restrictions related to cirrhosis could be withdrawn in the future [129]. In fact, it has been mentioned that acarbose could even be used to treat hepatic encephalopathy but there is no approved indication [130].

Indications for α-glucosidase inhibitors

Alpha-glucosidase inhibitors could be considered as monotherapy in the following cases:

1. Patients with normal or slightly elevated fasting blood glucose and postprandial hyperglycemia: In these patients, postprandial hyperglycemia contributes significantly to increased HbA1c [1]. This population includes most recently diagnosed patients. When diet and exercise fail to normalize blood glucose, AGI can bring glycemic control into the desirable range for an important portion of these patients.

2. Elderly diabetic patients: There is now evidence that this patient population significantly benefits from AGI treatment [25,30,58]. The absence of hypoglycemic risk is a major advantage in this subgroup and moderately decreased kidney function is not a contraindication.

3. As an alternative to other oral hypoglycemic agents when they are contraindicated: Alpha-glucosidase inhibitors can also be prescribed in combination with other hypoglycemic agents in various situations:
   a. as an adjunct to other hypoglycemic agents that have not produced adequate glycemic control. Combination of AGI
with all other treatments has demonstrated a moderate but significant, constant, and sustained glycemic improvement.

- as an alternative for patients not controlled with a combination of oral hypoglycemic agents and who refuse insulin.

The proper use of AGI requires attention on two major points:

1. To be effective, they must be administered at the onset of meals with the first bite and not later than 15 min after meals.
2. Therapy must be initiated with a low dose and titrated upward slowly based on PPG as well as gastrointestinal tolerance.

**Recommended doses are as follows:**

- 25–100 mg three times daily for acarbose;
- The maximum recommended dose for patients <60 kg is 50 mg three times daily
- The maximum recommended dose for patients >60 kg is 100 mg three times daily
- 25–100 mg three times daily for miglitol
- 0.2 mg three times daily for voglibose.

**Conclusions**

Alpha-glucosidase inhibitors are drugs that delay the digestion of complex carbohydrates by acting as competitive inhibitors of intestinal α-glucosidase enzymes that hydrolyze di-, oligo-, and polysaccharides into monosaccharides, an essential step for absorption. Thus, these drugs decrease the rise in postprandial plasma glucose and consequently insulin.

In subjects with IGT, AGI delay the onset of T2DM and hypertension, and possibly reduce macrovascular events. In type 2 diabetic patients, AGI have shown moderate but constant and sustained reductions of HbA1c (~0.7%) regardless of concomitant antidiabetic medication.

Additional benefits of AGI are the lack of hypoglycemia with monotherapy, no weight gain, and an excellent safety profile. Gastrointestinal symptoms are the consequence of undigested carbohydrates reaching the colon, where they are fermented by bacteria, thus causing flatulence, abdominal discomfort, or bloating. Hence the necessity of initiating treatment at a low dose and increasing the dosage slowly to minimize these side effects.

**References**


50 Pan C, Yang W, Barona JP, et al.: Comparison of vildagliptin and acarbose monotherapy in patients with Type 2 diabetes:


83 Seifarth C, Begemann J, Holst JI, et al.: Prolonged and enhanced secretion of glucagon-like peptide 1 (7-36 amide) after oral...


111 Acarbose Cardiovascular Evaluation (ACE) 2008: www.dtu.ox.ac.uk/ace/index (last accessed June 10, 2014).

CHAPTER 46

Combination therapy in type 2 diabetes mellitus

Muhammad Abdul-Ghani1, Ralph A. DeFronzo1 and Roy Eldor2
1Diabetes Division, Department of Medicine, The University of Texas Health Science Center at San Antonio (UTHSCSA), San Antonio, TX, USA
2LSD Clinical Research Diabetes and Endocrinology, Merck Research Laboratories, Rahway, NJ, USA

Key points

- Type 2 diabetes mellitus (T2DM) is characterized by multiple pathophysiologic abnormalities
- Multiple antidiabetic agents used in combination will be required to correct the multiple pathophysiologic disturbances and reduce/maintain the HbA1c \( \leq 6.5–7.0\% \)
- Medications that prevent the progressive beta cell failure (GLP-1 receptor agonists and thiazolidinediones) and improve the insulin resistance (thiazolidinediones) should be considered as first line agents
- Metformin and sulfonylureas are the most commonly prescribed combination worldwide but neither drug prevents/delays beta cell failure in T2DM and neither drug improves insulin sensitivity
- Metformin and thiazolidinediones have an additive effect to reduce the HbA1c and combination therapy does not cause hypoglycemia
- Combination therapy with metformin (which increases GLP-1) plus a DPP4 inhibitor has good efficacy in reducing HbA1c and does not cause hypoglycemia
- SGLT2 inhibitors work at all stages (early to late) of the natural history of T2DM and can be used in combination with all other antidiabetic agents, including insulin
- Triple combination therapy with pioglitazone/metformin/GLP-1 receptor agonist is especially effective in normalizing the HbA1c, even in T2DM patients with a starting HbA1c > 10.0%
- Intensive insulin therapy in combination with or followed by combination therapy with metformin and/or thiazolidinedione and or SGLT2 inhibitor, and/or GLP-1 receptor agonist can be effectively employed in poorly controlled, new-onset T2DM patients to normalize HbA1c

Natural history of type 2 diabetes

Individuals destined to develop T2DM inherit genes that make their tissues resistant to insulin [1–3,8–13]. Insulin resistance in liver results in overproduction of glucose during the fasting state [14] and impaired suppression of hepatic glucose production (HGP) by insulin [15], as occurs following a meal [16]. Insulin resistance in muscle [15,17–19] is manifest by impaired glucose uptake following carbohydrate ingestion, resulting in postprandial hyperglycemia [16]. Although origin of insulin resistance can be traced to its genetic background [2,8], the current diabetes epidemic is related to the obesity epidemic and physical inactivity [20]. Both obesity [21] and decreased physical activity [22] are insulin-resistant states. As long as \( \beta \) cells augment insulin secretion sufficiently to offset insulin resistance, glucose tolerance remains normal [1–3,23]. However, with time \( \beta \) cells begin to fail, and initially postprandial, and subsequently fasting plasma glucose levels...
Combination therapy in type 2 diabetes mellitus

Decreased incretin effect

HYPERGLYCEMIA

Decreased insulin secretion

Islet-α cell

Increased glucagon secretion

Increased HGP

Neurotransmitter dysfunction

Figure 46.1 The Ominous Octet: eight distinct pathophysiologic defects contribute to the pathogenesis of type 2 diabetes mellitus. Source: DeFronzo 2009 [3].

increase, leading to overt diabetes [1–3,5–7,24]. This natural history of T2DM [1–3] is depicted in Figure 46.2 [25].

Beta-cell function

The plasma insulin response to insulin resistance typically is increased early in the natural history of T2DM (Figure 46.2). However, plasma insulin response to glucose challenge does not provide a valid index of β-cell function [26]. β-Cells respond to an increment in glucose (ΔG) with an increment in insulin (ΔI) [26]; they also increase their secretion of insulin to compensate for the insulin resistance [23,26,27]. Thus, the gold standard for β-cell function is the insulin secretion/insulin resistance (disposition) index (ΔI/ΔG ÷ IR). Results of the SAM Study and VAGES have shown that subjects in the upper tertile of normal glucose tolerance have lost >50% of β-cell function, while subjects in the upper tertile of IGT have lost ~80% of β-cell function [1–7]. Similar conclusions are evident from other publications [10,25,28,29]. Further, Butler et al. [30] have demonstrated that, as individuals progress from NGT to impaired fasting glucose (IFG), there is a decline in β-cell mass which continues with progression to diabetes [31].

In summary, although insulin resistance in liver and muscle are well established early in the natural history of the disease, T2DM does not occur in the absence of progressive β-cell failure.

Insulin resistance

Liver. Following an overnight fast, liver of nondiabetic individuals produces glucose at ~2 mg kg⁻¹ per min [1–3,14]. In T2DM, the rate of basal HGP is increased, averaging ~2.5 mg kg⁻¹ per min [1–3,14]. This amounts to addition of 25–30 g of glucose to the systemic circulation every night and is responsible for the increase in FPG concentration. This hepatic overproduction of glucose occurs despite fasting plasma insulin levels that are increased two- to threefold, indicating severe hepatic resistance to insulin.

Muscle. Using the insulin clamp technique [32] with limb catheterization [1–3,15,17–19], it has conclusively demonstrated that lean, as well as obese, T2DM individuals are severely
resistant to insulin and that muscle is the primary site of insulin resistance. Multiple intramyocellular defects in insulin action are present in T2DM [1–3,33,34], including impaired glucose transport and phosphorylation [17], reduced glycogen synthesis [35], decreased glucose oxidation [5] and more proximal defects in insulin signal transduction [34,36].

Ominous Octet. In addition to the Triumvirate (β-cell failure; insulin resistance in muscle and liver) [1], at least five other pathophysiologic abnormalities contribute to glucose intolerance in T2DM [3] (Figure 46.1): (i) adipocyte resistance to antilipolytic effect of insulin leads to increased plasma FFA and elevated intracellular levels of toxic lipid metabolites in liver/muscle and β cells, causing insulin resistance and β-cell failure/apoptosis [15]; (ii) decreased incretin (GLP-1 and GIP) effect resulting from impaired GLP-1 secretion, but more importantly from severe β-cell resistance to both GLP-1 and GIP [37]; (iii) increased glucagon secretion by α cells and enhanced hepatic sensitivity to glucagon, leading to increased basal HGP and impaired HGP suppression by insulin [38,39]; (iv) enhanced renal glucose reabsorption by the SGLT2 transporter, contributing to maintenance of elevated plasma glucose levels [40]; (v) CNS resistance to the anorectic effect of insulin and altered neurosynaptic hormone secretion contributing to appetite dysregulation, weight gain, and insulin resistance in peripheral tissues and liver [41,42].

Implications for therapy

The preceding review of T2DM pathophysiology has important therapeutic implications: (i) effective treatment will require multiple drugs in combination to correct multiple pathophysiologic defects; (ii) treatment should be based upon reversal of known pathogenic abnormalities and NOT simply on reduction of A1c; (iii) therapy must be started early in the natural history of T2DM, if progressive β-cell failure is to be prevented.

Current therapeutic options as they relate to pathophysiologic derangements present in T2DM are depicted in Figure 46.3. In liver, both metformin [43–45] and thiazolidinediones (TZDs) [46–50] are potent insulin sensitizers and reduce the increased basal rate of hepatic glucose production. In muscle, TZDs are potent insulin sensitizers [46–48,51–53], whereas metformin is, at best, a weak insulin sensitizer [43,45,52]. Since TZDs work through the classic insulin signaling pathway [53,54], whereas metformin works through AMP kinase pathway [55,56], combination therapy with TZD plus metformin gives a completely additive effect to reduce HbA1c [57,60]. Further, hypoglycemia is not encountered because these drugs are insulin sensitizers. In adipose tissue, TZDs are excellent insulin sensitizers and potent inhibitors of lipolysis [61–63]. TZDs also effectively mobilize fat from muscle, liver, and β-cells, ameliorating lipotoxicity [47,62–67].

In the β cell, sulfonylureas, glinides, and incretin-based therapies (GLP-1 analogues and DPP-4 inhibitors) augment insulin secretion. However, only GLP-1 analogues [68–70] and TZDs [71,72] improve/preserve β-cell function and demonstrate durability of glycemic control [60,68,72–81]. Nonetheless, the two most commonly prescribed drugs worldwide are sulfonylureas and metformin, and neither exert any β-cell protective effect.

The GLP-1 analogue, exenatide, augments and preserves β-cell function for at least 3 years [68,69]. The β-cell protective effect of GLP-1 analogues has its onset within 24 h [70] and persists as long as GLP-1 therapy is continued [68,69,80]. Further, both exenatide and liraglutide inhibit the appetite center in CNS leading to weight loss [80–84], inhibit glucagon secretion [82–84], and delay gastric emptying [82–84]. By promoting weight loss and depleting lipid from muscle and liver, GLP-1 analogues also improve muscle and hepatic insulin sensitivity [68,69]. GLP-1 analogues correct multiple cardiovascular risk factors [80,85]. Although DPP-4 inhibitors share some characteristics with GLP-1 analogues, they do not raise plasma GLP-1 levels sufficiently to offset the β cell GLP-1 resistance [86,87]. Not
surprisingly, their ability to augment insulin secretion and reduce HbA1c are significantly less than GLP-1 analogues [86,87], and they do not promote weight loss [88]. The major mechanism of action of DPP-4 inhibitors to improve glycemic control is mediated via inhibition of glucagon secretion and subsequent decline in hepatic glucose production [89,90].

The recently approved sodium glucose transporter 2 (SGLT2) inhibitors demonstrate modest efficacy in reducing HbA1c, promote weight loss, reduce blood pressure, and can be added to all approved antidiabetic agents including insulin [40,91].

**Combination therapy**

In the present chapter we discuss combination therapy in two clinical situations: (i) initiation of combination therapy in newly diagnosed T2DM patients; (ii) add-on therapy in T2DM patients inadequately controlled with one or more oral agents. In the latter situation, we review use of basal insulin in T2DM patients suboptimally controlled with oral agents or GLP-1 analogues. Use of a mixed-split insulin regimen in this situation will not be addressed.

**Newly diagnosed T2DM patients**

When initiating combination therapy in newly diagnosed T2DM patients, the following considerations are of paramount importance:

1. **Therapy should be able to achieve the desired level of glycemic control:** HbA1c < 6.5% (EASD and AACE) [92] and ≤ 7.0% (ADA) [93]. However, we believe that, in newly diagnosed T2DM patients without cardiovascular disease, optimal glycemic control should be HbA1c < 6.0% and as close to normal as possible, while avoiding untoward adverse events, primarily hypoglycemia [3,94]. This is consistent with the expanded ADA statement [93].

2. **In the majority of newly diagnosed T2DM patients, monotherapy will not be sufficient to reduce HbA1c < 6.5–7.0%, or most optimally < 6.0%.** Therefore, combination therapy will be required [3].

3. **The medications used in combination therapy should have an additive effect and individual drugs should correct known pathophysiologic disturbances present in T2DM.** Otherwise, long-term durability of glycemic control will not be achieved.

4. **Progressive β-cell failure is responsible for the progressive HbA1c rise in T2DM [1–7].** Therefore, medications used to treat T2DM should preserve/improve β-cell function to produce a durable effect on glycemic control.

5. **Because underlying insulin resistance is a core defect in T2DM and accelerates β-cell failure, medications should also ameliorate insulin resistance in muscle and liver.**

6. **Because T2DM is a cardiovascular equivalent, it is desirable that drugs exert beneficial effects to improve cardiovascular (CV) risk factors and decrease CV events.**

7. **Obesity is a major problem in T2DM.** Therefore, it is desirable that combination therapy be weight neutral and, ideally, promote weight loss.

8. **Combination therapy should be safe.** No single antidiabetic agent can correct all of these pathophysiologic disturbances and multiple agents, used in combination, will be required to optimally control T2DM patients.

**Initial combination therapy in new onset diabetes**

**Rationale**

Although ADA/EASD recommend starting therapy with metformin as monotherapy [93], many considerations favor initiating combination therapy in newly diagnosed T2DM individuals [3]. Initiating therapy with multiple agents that have different mechanisms of action allows correction of more metabolic abnormalities than can be achieved with monotherapy. Further, the HbA1c decrease produced by a single antidiabetic agent, for example metformin, sulfonylurea, pioglitazone, GLP-1 analogue, is ~1.0–1.5% depending upon starting HbA1c. Thus, in newly diagnosed T2DM with HbA1c > 8.0–8.5%, it is unlikely that a single agent will achieve the recommended goal (HbA1c < 6.5–7.0%) and virtually no one will achieve HbA1c < 6.0%. When maximum dose metformin, sulfonylurea or thiazolidinedione is initiated as monotherapy, only 25–40% of newly diagnosed T2DM subjects achieve HbA1c < 6.5–7.0% [95,96], and an even smaller percentage achieve an acceptable HbA1c with acarbose or repaglinide [97]. Thus, most patients with HbA1c > 8.0–8.5% will require initial combination therapy. Moreover, because different agents lower plasma glucose via different mechanisms, combination therapy will have an additive or even synergistic effect to reduce HbA1c compared to each agent alone. Simultaneous improvements in β-cell dysfunction and insulin sensitivity are more likely to cause durable HbA1c reduction. Lastly, initiation of combination therapy allows use of submaximal doses of each antidiabetic agent and results in a lower adverse event rate. Consistent with this approach, AACE recommends starting newly diagnosed diabetic subjects with HbA1c > 8.0% on multiple antidiabetic agents [92]. A list of available combination tablets and their efficacy in reducing A1c is provided in Tables 46.1 and 46.2.

**Sulfonylurea plus metformin**

The most commonly used combination therapy worldwide is metformin plus sulfonylurea. Because the glucose-lowering mechanisms with metformin (inhibition of hepatic glucose production) and sulfonylureas (enhanced insulin secretion) differ, combined sulfonylurea/metformin therapy has an additive effect to reduce HbA1c versus either agent alone [44,95,96]. Sulfonylurea addition to metformin, metformin addition to sulfonylurea and initiation of combination metformin/sulfonylurea
therapy in drug-naïve T2DM subjects lowers HbA1c by ~2.0%. Herman et al. [95] randomized 144 T2DM subjects to metformin alone, glyburide alone or combination therapy. The decrease in HbA1c in subjects receiving metformin alone and glyburide alone was 0.9% and 1.3%, and only 66% and 62%, respectively, achieved the treatment goal. The decrease in HbA1c in subjects started on combination therapy was 2.2%. Garber et al. [96] randomized 806 drug-naïve T2DM patients to glyburide (2.5 mg), metformin (500 mg), glyburide/metformin (1.25/250 mg), or glyburide/metformin (2.5/500 mg). Patients receiving combination glyburide/metformin therapy (both low and high doses) had significantly greater HbA1c reductions (1.48% and 1.53%, respectively) than subjects receiving metformin alone (1.03%) or glyburide alone (1.24%). Although the HbA1c decrease was comparable in both combination therapy groups, hypoglycemia was threefold more common in the high-versus low-dose combination group (16% vs. 5%). Moreover, despite greater HbA1c decrease with glyburide/metformin (1.25/250 mg) versus glyburide alone, hypoglycemic events were slightly lower in the combination group (5.1% vs. 6.3%). These results demonstrate that initial combination therapy with metformin/sulfonylurea has an additive effect to reduce HbA1c and allows use of submaximal doses of each drug, minimizing hypoglycemia and weight gain.

Amongst sulfonylurea/metformin combination tablets, Glucovance® (metformin/glyburide) has achieved the most widespread usage. In drug-naïve T2DM patients (HbA1c = 8.2%), Glucovance reduced HbA1c by 1.5%, and this was 0.5% more than when glyburide and metformin were administered as separate tablets. The greater HbA1c decrement in Glucovance-treated patients resulted from an augmented insulin response from increased glyburide bioavailability. Consistent with this, T2DM patients on chronic treatment with glyburide and metformin as separate tablets, experienced an additional 0.6% HbA1c reduction when switched to Glucovance [99]. Metaglip™ (metformin/glipizide) is slightly less potent than Glucovance in reducing HbA1c, most likely because Metaglip formulation does not result in increased glipizide bioavailability.
A major limitation of combination therapy with sulfonylurea/metformin is that neither agent preserves β-cell function or improves insulin sensitivity in muscle. UKPDS demonstrated that, despite an additional HbA1c decrease when metformin was added to glyburide or glyburide to metformin, HbA1c progressively increased and after 10 years the majority (∼60%) of subjects required insulin therapy [20,61].

**Metformin plus repaglinide/nateglinide**

Repaglinide and nateglinide are short-acting insulin secretagogues, given 2–3 times daily with meals. Because metformin primarily reduces FPG by suppressing HGP, while nateglinide targets postprandial hyperglycemia, combination metformin/nateglinide therapy improves both fasting and postprandial hyperglycemia. Three large clinical trials [97,98,100] have demonstrated that combined metformin/nateglinide therapy safely and effectively lowers HbA1c in drug-naïve T2DM patients. In 701 drug-naïve T2DM individuals (HbA1c ∼8.3%) nateglinide (120 mg) plus metformin (500 mg) versus nateglinide alone, metformin alone, and placebo decreased HbA1c by −1.4%, −0.5%, −0.8%, and +0.5%, respectively [97]. Seventy percent of subjects who received combination therapy achieved HbA1c < 7.0% versus only 34%, 41%, and 17% who received nateglinide, metformin, and placebo, respectively. Metformin had greater efficacy in lowering FPG, while nateglinide was more effective in reducing postprandial glucose levels. There were more hypoglycemic events in subjects receiving combination therapy (26%) versus nateglinide (12%) or metformin (10%) alone. However, most of these events were mild.

Two studies [98,100] comparing long-term efficacy of initial combination therapy with metformin/nateglinide versus metformin/sulfonylurea in drug-naïve T2DM individuals reported conflicting results. In PRESAVE-Beta [98] 404 drug-naïve T2DM subjects were randomized to glyburide (10 mg)/metformin (2000 mg) or nateglinide (120 mg with meals)/metformin (2000 mg). HbA1c decrease at 104 weeks was similar in both groups (1.6% in glyburide/metformin vs. −1.4 in nateglinide/metformin), and the number of subjects maintaining HbA1c < 7.0% at 2 years was similar in both groups (39% vs. 43%). Glyburide/metformin was more effective in decreasing FPG, while nateglitide/metformin was more effective in lowering postprandial glucose. Despite the similar HbA1c decrease, hypoglycemia was more common with glyburide/metformin (17.7%) versus nateglinide/metformin (8.2%). In a 52-week study Derosa et al. [100] compared the efficacy of initial combination therapy with nateglinide/metformin versus glibenclamide/metformin in 248 drug-naïve T2DM individuals. Unlike PRESAVE-Beta, the HbA1c decrease was greater with nateglinide/metformin (from 8.1% to 6.4%) compared to glibenclamide/metformin (1.7% vs. 10.9%).

The efficacy of repaglinide (1–4 mg per meal)/metformin versus nateglinide (120 mg with meals)/metformin was examined in a 20-week open-label, parallel-group, randomized trial in 192 T2DM patients. Repaglinide/metformin produced greater HbA1c (1.3% vs. 0.7%, $p < 0.01$) and FPG (39 versus 21 mg dL$^{-1}$, $p < 0.01$) reductions compared to nateglinide/metformin. Paradoxically, the change in plasma glucose, glucagon and insulin concentrations AUCs after a test meal were comparable in both groups.

In summary, initial combination therapy with metformin/nateglinide and metformin/repaglinide is effective in lowering HbA1c with a lower rate of hypoglycemia compared to metformin/sulfonylurea therapy. However, long-term durability (≥2 years) of this combination has never been tested.

**Metformin plus thiazolidinediones**

Metformin inhibits hepatic gluconeogenesis via activation of AMP kinase [55,56], while thiazolidinediones have both direct (PPAR activation) and indirect (plasma FFA reduction, depletion of intramyocellular and intrahepatic fat, increased adiponectin) effects to improve insulin sensitivity in muscle and liver [18,43–45,83]. TZDs also enhance/preserve β-cell function [71]. Thus, combined metformin/thiazolidinedione therapy has an additive HbA1c lowering effect with low incidence of hypoglycemia [59,60]. Metformin also reduced the MACE endpoint of myocardial infarction, stroke, and cardiovascular death in UKPDS [29], while pioglitazone corrects multiple components of insulin resistance syndrome [46,51], reduces carotid intimal thickness [73], prevents restenosis of coronary artery stents [101], and decreases coronary atheroma volume [74]. Pioglitazone also decreased the second principal endpoint (composite of myocardial infarction, stroke, death) in PROACTIVE [102]. Rosiglitazone, unlike pioglitazone, has been associated with increased incidence of cardiovascular events [27] and withdrawn from most markets worldwide.

Multiple clinical studies have demonstrated the additive metabolic benefit of TZDs plus metformin. In 600 drug-naïve T2DM patients (HbA1c ∼8.7%) treated with pioglitazone (30 mg/metformin 1700 mg) the HbA1c decrement was twofold greater than in subjects who received pioglitazone alone (−0.96%) or metformin alone (−0.99%) and 64% of subjects achieved HbA1c < 7.0% [103]. The HbA1c decrease was accompanied by a greater decrease in FPG in the combination group (−40 mg dL$^{-1}$) compared to each therapy alone. Despite greater decreases in HbA1c and FPG with combination therapy, hypoglycemia was similar to the monotherapy groups. Subjects receiving combination pioglitazone/metformin experienced improvement in cardiovascular risk factors: increased HDL cholesterol, decreased plasma triglyceride, decreased small dense LDL particle number, decreased hsCRP, and increased plasma adiponectin. However, similar changes in cardiovascular risk factors were observed in subjects treated with pioglitazone alone.

Similar results were reported by Rosenstock et al. [104] using initial combination therapy with rosiglitazone (8 mg)/metformin (2000 mg) versus metformin alone and rosiglitazone alone in 468 drug-naïve T2DM patients (HbA1c = 8.8%). After 32 weeks, the HbA1c decrease was significantly greater in
subjects receiving rosiglitazone/metformin (−2.3%) versus metformin alone (−1.81%) and rosiglitazone alone (−1.6%). Rosiglitazone/metformin did not increase risk of hypoglycemia compared to either monotherapy and produced greater decrease in hsCRP and increase in plasma adiponectin concentration, which did not differ from changes observed with rosiglitazone alone.

In summary, initial combination therapy with pioglitazone/metformin offers multiple advantages: excellent glycemic control, improved cardiovascular risk profile, regression of established coronary artery atheroma volume, decreased cardiovascular events (PROactive), correction of core T2DM pathophysiologic defects (insulin resistance and impaired β-cell function), and low incidence of hypoglycemia.

**Metformin plus DPP-4 inhibitor**

DPP-4 inhibitors lower HbA1c by increasing plasma GLP-1/GIP levels, leading to increased glucose-stimulated insulin secretion and inhibition of glucagon secretion [89,90]. Metformin also increases plasma GLP-1 levels [105]. Thus, the combination of metformin (reduces HGP) plus DPP-4 inhibitor (insulin secretagogue) would be expected to produce an additive effect to reduce HbA1c [106]. In 1091 drug-naïve T2DM patients combined therapy with sitagliptin (100 mg)/metformin (1000 mg and 2000 mg) was compared to sitagliptin alone. In subjects receiving sitagliptin monotherapy, HbA1c decreased by −0.66% versus −0.82% in subjects receiving 1000 mg day−1 metformin. However, the HbA1c decrease in subjects receiving sitagliptin/metformin 1000 mg day−1 was −1.4% and sitagliptin/metformin 2000 mg was −1.9 versus 0.17% increase in placebo group (Figure 46.4). Both metformin doses were more effective than sitagliptin in lowering FPG. Combination of sitagliptin with either metformin dose produced additive decreases in FPG, 2h-PG, and glucose AUC versus either therapy alone (Figure 46.4). The rate of hypoglycemia was low (<2.5%) in both combination therapy groups and did not differ from that of either therapy alone. During a 102-week extension period, subjects who received initial sitagliptin/metformin therapy continued to experience greater HbA1c reduction versus either monotherapy. However, the magnitude of HbA1c decrease began to wane at week 24, and statistical testing on durability of HbA1c control was not performed in the extension study because of high dropout; only ∼70% of subjects with HbA1c < 7.0% at week 24 maintained this target at week 102, indicating progressive loss of efficacy [107]. This observation suggests that DPP-4 inhibitors do not alter the progressive β-cell failure that is in T2DM.

Initial vildagliptin/metformin combination therapy in drug-naïve T2DM subjects (HbA1c ∼ 8.7%) reduced HbA1c by 1.8, 1.6, 1.1, and 1.4%, respectively, in subjects receiving sitagliptin/metformin (2000 mg), sitagliptin/metformin (1000 mg), sitagliptin alone, and metformin alone (p < 0.01 combination vs. monotherapy) [108]; 65% of subjects receiving combination therapy achieved HbA1c < 7.0%. Combination therapy produced a greater FPG reduction without increased hypoglycemic risk versus monotherapy.

In summary, combination DPP-4 inhibitor/metformin therapy produces an additive HbA1c reduction compared to each monotherapy. The magnitude of HbA1c decrease is smaller than produced by combined sulfonylurea/metformin therapy. However, unlike combined sulfonylurea/metformin therapy, DPP-4 inhibitor/metformin treatment is not associated with increased hypoglycemic risk. Available 2-year data indicate that HbA1c lowering of DPP-4 inhibitor/metformin is similar to sulfonylurea/metformin.

**Thiazolidinedione plus DPP-4 inhibitor**

Four studies with different DPP-4 inhibitors (sitagliptin, vildagliptin, alogliptin, linagliptin) plus pioglitazone have demonstrated that combination therapy produces greater HbA1c reduction versus either agent alone in drug-naïve T2DM subjects. Initial combination therapy with vildagliptin 100 mg plus two doses of pioglitazone (15 and 30 mg) in 607 drug-naïve T2DM individuals (HbA1c ∼ 8.7%) reduced HbA1c by 1.7% (Vilda/Pio 15 mg), 1.9% (Vilda/Pio 30 mg), 1.1% (Vilda) and 1.4% (Pio) [109]. More subjects receiving vildagliptin/pioglitazone 30 mg achieved HbA1c < 7.0% (65%) versus vildagliptin alone (42%) or pioglitazone 30 mg alone (43%). Pioglitazone monotherapy was more effective than vildagliptin in lowering FPG, while vildagliptin had greater effect on postprandial glucose. Combined pioglitazone 30 mg/vildagliptin produced greater reductions in fasting and postprandial glucose versus either therapy alone. During meal tolerance test, vildagliptin alone and vildagliptin/pioglitazone produced greater increase in the ratio of plasma insulin AUC/plasma glucose AUC than pioglitazone alone. Although these results demonstrate improved β-cell function by vildagliptin, they do not exclude a beneficial effect of pioglitazone on β-cell function, since change in insulin secretion was
not factored by the severity of insulin resistance (IR). Since pioglitazone is a powerful insulin sensitizer [82,83], the effect of combined pioglitazone/vildagliptin therapy on β-cell function \((ΔΙ/ΔG \div IR)\) is much greater than the ratio between the plasma insulin AUC/plasma glucose AUC. Weight gain was more common with pioglitazone versus vildagliptin monotherapy and more common with high-dose pioglitazone versus low-dose pioglitazone combination therapy. Despite greater weight gain with both pioglitazone combinations, the TZD caused a greater HbA1c decrement. Incidence of hypoglycemia was low (<1%) in combined pioglitazone/vildagliptin groups and did not differ from that in each monotherapy group.

A similar additive HbA1c lowering effect was reported in initial combination therapy with sitagliptin 100 mg/pioglitazone 30 mg compared to pioglitazone alone (~2.4% vs. ~1.5%) over 54 weeks [110]. The greater HbA1c decrease with sitagliptin/pioglitazone reported by Yoon et al. [109] versus Rosenstock et al. [109] with vildagliptin/pioglitazone is explained by the higher starting HbA1c in the former study (9.5% vs. 8.7%).

Combined linagliptin/pioglitazone therapy in drug-naïve T2DM individuals reduced HbA1c by ~1.06% versus ~0.50% in the pioglitazone-treated group [111]. The initial HbA1c in this study was comparable to that in the Rosenstock study (8.6 vs. 8.7%) and cannot account for the smaller HbA1c decrease with combined linagliptin/pioglitazone, but could be explained by the higher Asian population (30%) in the Yoon study [110]. Increased efficacy of incretin-based therapy (DPP-4 inhibitors and GLP-1 analogues) has been reported in Asians versus other ethnic groups [112].

Combined alogliptin/pioglitazone therapy also produced an additive effect to reduce HbA1c in 655 drug-naïve T2DM patients (HbA1c ~ 8.8%) randomized to receive 30 mg pioglitazone, 25 mg alogliptin, 12.5 mg alogliptin plus 30 mg pioglitazone, or 25 mg alogliptin plus 30 mg pioglitazone [113]. The HbA1c decrease was ~1.15% (Pio), ~0.96% (Alo), ~1.56% (Pio + Alo 12.5 mg) and ~1.71% (Pio + Alo 25 mg). The FPG decrease followed the same pattern as HbA1c and was greater in combination therapy groups versus monotherapy groups. Hypoglycemia incidence was low (<3%) and similar in all four groups.

In summary, combination therapy with DPP-4 inhibitor/pioglitazone produces an additive effect on HbA1c versus each therapy alone without increased hypoglycemic risk. Although neither pioglitazone nor DPP-4 inhibitors are recommended by ADA/EASD as initial therapy in T2DM, subjects who cannot take metformin (gastrointestinal side effects or impaired renal function) can be treated with this combination.

**Combination of SGLT2 inhibitors plus metformin**

Inhibitors of renal sodium glucose transporter 2 (SGLT2 inhibitors) are a novel class of antidiabetic agents recently approved in Europe and the US. Because of their unique mechanism of action to inhibit reabsorption of filtered glucose [40,91], they are equally effective in individuals with severe insulin resistance and/or severe β-cell failure, as well as in newly diagnosed T2DM [114]. Combination therapy with dapagliflozin (5 and 10 mg) plus metformin 2000 mg produced additive decreases in HbA1c with both doses compared to each monotherapy [115]. Dapagliflozin 5 mg/metformin decreased HbA1c by ~2.05% from baseline HbA1c = 9.21% compared to decreases of ~1.19% and ~1.30% with dapagliflozin alone or metformin alone. Similar results were observed with dapagliflozin 10 mg plus metformin 2000 mg. The HbA1c decrease with dapagliflozin 10 mg was similar to that with dapagliflozin 5 mg, indicating dapagliflozin 5 mg produced a maximal reduction in HbA1c. The FPG decrease with combination dapagliflozin/metformin therapy was additive compared to each therapy alone. Approximately half of subjects receiving combination therapy with both doses of dapagliflozin/metformin reached HbA1c < 7.0% versus 35% and 25–30% with metformin and dapagliflozin monotherapy. Incidence of hypoglycemia was low (<3.0%) and similar in combination and monotherapy groups. Subjects in both combination therapy groups experienced ~3% decrease in body weight over 24 weeks. Adverse events included small increases in urinary tract infection (~5%), vulvovaginitis (~10%), and balanitis (~15%) in subjects receiving dapagliflozin.

**Initial triple combination therapy**

Initial triple combination therapy with sulfonylurea/metformin/pioglitazone was examined in 58 newly diagnosed T2DM patients [116]. All subjects received lead-in treatment for 3 months with metformin plus insulin. At 3 months, subjects were randomized to triple oral therapy with pioglitazone/metformin/glyburide \((n=29)\) or continued with metformin/insulin \((n=29)\). After the 3-month lead-in period, HbA1c decreased from 10.8% to 5.9% and during 3 years of follow-up subjects in both groups maintained the initial HbA1c reduction (6.1% and 6.0%, respectively). Although demonstrating maintenance of excellent glycemic control for 3 years with metformin/insulin or metformin/sulfonylurea/pioglitazone, the insulin dose had to be increased from 64 to 80 units per day at study end, indicating that β-cell failure continued in this group. Moreover, good glycemic control was achieved at the expense of a relatively high rate of hypoglycemia and weight gain. Insulin-treated subjects experienced 8.1 hypoglycemic events per year versus 6.1 events per year in the triple therapy group, and 55/58 study participants experienced at least one hypoglycemic event. Weight gain was greater (10.1 kg) in the triple therapy group versus the insulin/metformin group (3.3 kg).

In an ongoing study we compared initial triple therapy with pioglitazone/metformin/exenatide twice daily with the “standard” ADA approach of metformin, then add sulfonylurea, then add basal insulin in 147 newly diagnosed T2DM patients with HbA1c = 8.6% [117]. After 2 years HbA1c reduction was greater with triple therapy versus the sequential ADA approach (2.5% vs. 2.0%, \(p < 0.01\)), triple therapy subjects lost 1.2 kg versus weight gain of 3.6 kg with the ADA approach, and hypoglycemia
incidence was 13.6-fold higher in the ADA-treated group. These results indicate that a combination approach focused on reversing underlying insulin resistance and β-cell dysfunction is superior and safer than an intervention of sequential therapy with metformin → sulfonylurea → insulin that does not reverse underlying pathophysiologic defects in T2DM.

**Combination therapy in T2DM patients inadequately controlled on antidiabetic therapy**

**Introduction**

In this section we consider add-on combination options for T2DM patients suboptimally controlled on single or multiple oral agents. Definition of suboptimal control will vary based upon associated medical disorders, especially cardiovascular disease (CVD), patient’s age, and diabetes duration. In elderly T2DM patients with long diabetes duration and clinically significant CVD, it would be prudent to choose a higher HbA1c goal, 7.0–7.5%, and avoid overly aggressive therapies that might induce hypoglycemia [118–121]. Three large prospective studies failed to demonstrate benefit of intensive glycemic control in this population. In contrast, younger patients with shorter disease duration and without CVD can benefit from more intensive glycemic control: A1c <7.0% (ADA), <6.5% (EASD), 5.5–6.0% (optimal).

Because metformin is the recommended drug of choice for initiation of therapy in most guidelines [92,122], we will focus on add-on therapy to metformin. However, in agreement with AACE [92], other agents (pioglitazone, GLP-1 analogues, and to a lesser extent DPP-4 inhibitors) are equally as effective as metformin in reducing A1c and have a more durable action because they improve β-cell function. Therefore, we also will discuss these alternative options.

Lastly, many T2DM individuals experience loss of β-cell function despite multiple oral antidiabetic agents and require addition of basal insulin to achieve glycemic control. However, our personal belief, based upon extensive clinical experience, is that long-acting insulin should not be added unless the subject has failed on a GLP-1 receptor agonist. Insulin therapy is associated with weight gain and hypoglycemia, requires dose titration, timing with meals, and frequent glucose monitoring, and is patient unfriendly. In contrast, GLP-1 analogues do not cause hypoglycemia, promote weight loss, do not require dose titration or intensive glucose monitoring, and are more patient friendly.

**Sulfonylureas/glinides: the treat-to-fail**

**approach**

Until recently, sulfonylureas have been considered the preferred drug of choice for add-on therapy to metformin in suboptimally controlled T2DM patients. This is attributed to their low cost and rapid onset of hypoglycemic effect. However, they lack “glycemic durability” and, within 1–2 years of treatment initiation, they lose their efficacy, resulting in a steady HbA1c rise to or above pretreatment levels [72] (Figure 46.5). Long-term studies examining durability of glycemic control with glinides are not available. However, nateglinide failed to prevent patients with impaired glucose tolerance from progressing to overt T2DM [123]. In a 2-year study in newly diagnosed T2DM subjects, durability of netaglinide plus metformin was comparable to glyburide plus metformin [123] and both groups experienced progressive HbA1c rise after the first year. Since deterioration of glycemic control is largely accounted for by a progressive decline in β-cell function [3,30], it is clear that both sulfonylureas and glinides fail to prevent the progressive β-cell failure [3,30] that is characteristic of T2DM. Consistent with this, *in vitro* studies have demonstrated a pro-apoptotic β-cell effect of sulfonylureas and glinides [124–128].

UKPDS clearly demonstrated that sulfonylureas had no β-cell protective effect in newly diagnosed T2DM patients [129]. After an initial drop in HbA1c, sulfonylurea-treated patients experienced progressive deterioration in glycemic control that paralleled the HbA1c rise in conventionally treated patients (Figure 46.5). In UKPDS sulfonylureas had no significant protective effect against atherosclerotic cardiovascular complications [24] and some studies have suggested that sulfonylureas may even accelerate atherogenesis [130,131]. Metformin-treated patients in UKPDS, after an initial decline in A1c, also experienced progressive deterioration in glycemic control (Figure 46.5) [132]. Using HOMA-β, UKPDS demonstrated that the relentless HbA1c rise observed with sulfonylureas and metformin resulted from progressive β-cell failure and within 3–5 years ~50% of diabetic patients required an additional pharmacologic agent to maintain HbA1c <7.0% [133–138]. Although some *in vitro* evidence suggests metformin may improve β-cell function [139], *in vivo* data from UKPDS fail to support a role for metformin in preservation of
β-cell function in humans. Importantly, however, metformin reduced macrovascular events in UKPDS [132], although by today’s standards the number of diabetic subjects in the metformin arm (n = 342) would be inadequate to justify any conclusions about cardiovascular protection.

UKPDS was designed as a monotherapy study. After 3–5 years, it became evident that neither metformin nor sulfonylurea monotherapy could prevent progressive β-cell failure and stabilize HbA1c at its starting level (133–137). Therefore, the study protocol was altered to allow metformin addition to the sulfonylurea arm and sulfonylurea addition to metformin [133–137]. Although addition of the second agent temporarily improved glycemic control, progressive β-cell failure continued and HbA1c rose progressively.

Multiple long-term (>1.5 years), active-comparator or placebo-controlled studies have examined the ability of sulfonylureas to produce durable HbA1c reduction in T2DM. Each study [60,73,75,123,140–143] showed that, after an initial decline in HbA1c, a variety of sulfonylureas (glyburide, glimepiride, and gliclazide) were associated with progressive β-cell failure with loss of glycemic control (Figure 46.6). There are no exceptions to this consistent loss of glycemic control with sulfonylureas after the initial 18 months of therapy. Thus, evidence-based medicine documents that the glucose-lowering effect of sulfonylureas is not durable and is associated with progressive β-cell failure [60,73,75,123,133–137,140–144].

Sulfonylurea treatment is associated with significant weight gain [145] and hypoglycemia [19,129]. No study has clearly implicated sulfonylureas/glinides as having a negative effect on cardiovascular events [146,147]. However, glibenclamide (glyburide) impairs ischemic preconditioning [49,148], and a possible association between sulfonylureas and adverse cardiovascular outcomes has been suggested [149–152]. However, other studies with gliclazide and nateglinide failed to find such an association [120,123,153–155]. Amongst β-cell secretagogues, shorter acting glinides and gliclazide are associated with reduced hypoglycemia compared to longer acting sulfonylureas such as glibenclamide [156–158].

In conclusion, we believe that sulfonylureas and glinides represent a poor option as add-on therapy to metformin, because they are associated with weight gain, hypoglycemia, progressive β-cell failure, and deterioration of glycemic control. The rapid, but short-lived, hypoglycemic effect is the primary advantage of this class and sulfonylureas can be used to initiate a short course of intensive glycemic control before instituting a more durable treatment plan. It should be recognized, however, that in many countries other newer antidiabetic agents (pioglitazone, GLP-1 analogues, DPP-4 inhibitors) are not available and/or are very expensive. In such circumstances, sulfonylureas may be the only option.

**Pioglitazone: unique benefits, unique side effects**

Rosiglitazone has been removed from the worldwide market because of cardiovascular safety concerns [159], and pioglitazone is the only representative of the TZD class. Pioglitazone is unique in that it exerts both a β-cell protective effect (71) and is a powerful insulin sensitizer in muscle and liver [46,51,160]. Not surprisingly, it has a durable effect of reducing HbA1c [123,141,161] with low risk of hypoglycemia [162].

Eight long-term (>1.5 years) active comparator or double-blind, placebo-controlled studies with TZDs (Figure 46.7) [60,73–75,123,141–143] have demonstrated that, after an initial decline in A1c, durable glycemic control is maintained because of preserved β-cell function in T2DM. In addition, there are six studies which demonstrate that TZDs prevent progression of IGT to T2DM [163–168]. These studies demonstrate that, in addition to their insulin sensitizing effect, TZDs had a major action to preserve β-cell function. In ACT NOW,

![Figure 46.6](image-url)  
**Figure 46.6** Long-term (>1.5 years) studies examining the effect of sulfonylurea therapy on A1c. See text for a more detailed discussion. Source: DeFronzo 2009 [3].
marked improvement in insulin secretion/insulin resistance (disposition) index (gold standard measure of β-cell function) was demonstrated with OGTT and FSIVGTT [168]. Similar results have been demonstrated in TRIPOD (troglitazone) was demonstrated with OGTT and FSIVGTT [168]. Similar greater are the improvements in insulin sensitivity and greater the weight gain, the greater is the A1c decline, and the

Pioglitazone causes fat weight gain [60], but the combined effects associated with pioglitazone therapy. A small increase in bone fractures, primarily related to trauma, has been documented in postmenopausal females [191], but not in premenopausal females or males and can be avoided by not using pioglitazone in this population, especially if bone mineral content is decreased. Although a slight increase in bladder cancer [192,193] has been reported in some studies, 8-year follow-up of a large prospective FDA-mandated study demonstrated a hazard ratio of 0.98 in pioglitazone versus non-pioglitazone-treated diabetics [194]. Lastly, a starting pioglitazone dose of 15 mg day⁻¹ titrated to 30 mg day⁻¹, provides 70–80% of glycemic efficacy with minimal side effects. Titrating to 45 mg day⁻¹ provides only a modest further increase in efficacy and a marked increase in side effects.

Pioglitazone/metformin combination therapy offers an effective, durable therapy that retards β-cell failure with little hypoglycemic risk. In a 24-week trial, pioglitazone/metformin reduced A1c by 1.8% (baseline A1c 8.6%), which was greater than the 1.0% reduction observed with metformin alone or pioglitazone alone [195]. As add-on to failing metformin therapy, pioglitazone reduced A1c by 0.9% over 2 years [60]. This combination has added value by improving disease parameters associated with NASH or NAFLD [66,176,177,196–199].

Combining pioglitazone with insulin effectively reduces A1c, but may cause excessive fluid retention and CHF [187]. Combining pioglitazone with SGLT-2 inhibitor may reduce peripheral edema [200] due to the diuretic effect of the SGLT2 inhibitor. Combining pioglitazone with GLP-1 analogue offsets the fat weight gain associated with pioglitazone monotherapy [201], and we particularly like this combination with or without metformin to institute therapy in newly diagnosed T2DM patients [3] (see subsequent discussion).
Combination therapy in type 2 diabetes mellitus

697

Decreased incretin effect

HYPERGLYCEMIA

Increased HGP

Islet-α cell

Increased glucagon secretion

TZDs

GLP1

Decreased glucose uptake

Increased glucose reabsorption

TZDs

GLP1

GLP1

Figure 46.8 Pathophysiologic disturbances reversed by triple combination therapy in type 2 diabetic patients treated with metformin plus pioglitazone plus a GLP-1 analogue.

**Metformin + GLP-1 analogue + pioglitazone: the pathophysiologic option**

Combination therapy with a biguanide, thiazolidinedione, and GLP-1 analogue offers a rational therapeutic approach targeting multiple pathophysiologic abnormalities in T2DM: muscle insulin resistance (pioglitazone), adipocyte insulin resistance (pioglitazone), pancreatic β-cell failure (GLP-1 analogue and pioglitazone), hepatic insulin resistance (metformin and pioglitazone), and excessive glucagon secretion (GLP-1 analogue) [3] with the added advantage of weight loss (GLP1 analogue) [209]. In a 20-week study of 73 T2DM patients treated with rosiglitazone, exenatide, or both (as add-on to metformin), significant improvements in β-cell function and insulin sensitivity were noted, with weight loss in all exenatide-treated groups [209]. In 233 T2DM patients treated with a thiazolidinedione plus metformin addition of exenatide reduced HbA1c by 1.0% and body weight by 1.5 kg [201]. Similar results have been reported by Liutkus et al. [221] and liraglutide addition to T2DM patients suboptimally controlled on metformin/rosiglitazone reduced HbA1c by 1.5% [222]. Newer once-weekly GLP-1 preparations provide all of the physiologic benefits of exenatide and liraglutide and facilitate patient adherence [223]. In poorly controlled T2DM patients, Bydureon® addition to metformin decreased A1c by 1.5% and caused 2.3 kg weight loss [205].

**Metformin + basal insulin: reliable but old fashioned**

Multiple studies have demonstrated that addition of basal insulin, glargine [224–229], and detemir [230,231] to metformin reduces A1c by 1.1–1.7%. However, this combination is associated with weight gain and hypoglycemia and does not correct any pathophysiologic abnormality present in T2DM [227,229–231]. It also introduces the need for self blood glucose monitoring to optimally control glycemia [232,233].

**Metformin + exenatide + basal insulin: the complementary combination**

The rationale for this combination is based upon the complementary effects of exenatide to reduce postprandial glucose and basal insulin to reduce FPG. Reduction in postprandial glucose results from exenatide’s effect to delay gastric emptying and inhibit HGP secondary to stimulation of insulin and inhibition of glucagon secretion [234]. The primary effect of basal insulin is to inhibit HGP during sleeping hours, leading to a reduction in FPG [228,235,236].

Combination therapy of basal insulin with exenatide or liraglutide has been approved in the US. In poorly controlled T2DM patients on insulin plus oral agents, exenatide reduced HbA1c (−0.87%), body weight (−5.2 kg), premeal and basal insulin dose, and sulfonylurea use [219]. In a short-term study, exenatide addition to insulin glargine plus metformin significantly reduced postprandial glucose excursion and HbA1c [237]. In a 30-week prospective study, addition of twice-daily exenatide to basal insulin reduced HbA1c by −0.69% and weight by −2.7 kg without increased hypoglycemia [238]. Combined liraglutide/insulin therapy in poorly controlled T2DM patients on metformin decreased HbA1c by 0.5%, primarily from a decrease in FPG [239]. In a survey based on data from the ABCD database, combination insulin/exenatide therapy effectively reduced A1c but was associated with a slightly higher rate of hypoglycemia (8.9 vs. 6.1%, p < 0.001) [240]. Improved glycemic control and weight loss have been reported in a 24-month retrospective analysis of poorly controlled T2DM patients following initiation of exenatide/glargine [241] and in a retrospective review of national US insurance claims database [242].
DPP-4 inhibitors: weak but easy

Dipeptidyl peptidase-4 inhibitors (DPP-4i) have gained widespread use as add-on therapy to metformin [243] because of weight neutrality, modest efficacy, and safety [244–249]. DPP-4i have a weak effect to augment insulin secretion [87,90,250]; their major effect to improve glycemic control is related to inhibition of glucagon secretion and resultant suppression of HGP [90]. In T2DM metformin increases endogenous GLP-1 secretion [105,251,252]. Combined metformin/DPP-4i therapy increases GLP-1 levels [253] and has an additive, even synergistic glucose lowering effect [247,254]. Compared with GLP-1 analogues, DPP-4i are clearly less effective. In a head-to-head cross-over study, exenatide was markedly superior to DPP-4i in reducing postprandial glucose, augmenting insulin and inhibiting glucagon secretion, and promoting weight loss [87]. In a head-to-head study, liraglutide reduced A1c by 0.6% more than sitagliptin [255]. Other studies have demonstrated superior efficacy of GLP-1 analogues in reducing A1c [256,257], augmenting β-cell function [258], and promoting weight loss [259]. A major advantage of DPP-4i is their ease of administration and excellent safety profile.

In combination with sulfonylureas, increased hypoglycemia has been observed with DPP-4i [244,260]. Because sulfonylureas have no effect to preserve β-cell function and cause hypoglycemia when used with DPP-4i, this combination is not recommended. In triple combination with metformin/pioglitazone (30 mg), the DPP-4ialogliptinresultedinbetter glycemic control and fewer pioglitazone dose-dependent side effects (edema and anemia) compared to metformin plus a higher dose of pioglitazone (45 mg) [261]. This study, as well as triple combination studies with metformin/pioglitazone/GLP-1 analogue [222,262], present a compelling new approach to diabetes therapy, similar to the hypertension treatment paradigm in which submaximal doses of therapeutic combinations result in safer, more robust effects.

DPP-4i plus metformin provide an attractive option for the geriatric diabetic population which is at increased risk of drug-related side effects, especially hypoglycemia. All published studies with DPP-4i have shown them to be as, if not more, effective in reducing A1c in elderly versus younger T2DM patients with an excellent safety profile [263–266]. DPP-4 inhibitors also appear to be especially effective in Asian populations.

Basal insulin analogues: when the β cells fail …

Insulin is the most potent antihyperglycemic available and, in theory, should be able to achieve normoglycemia in all T2DM patients [267,268]. Insulin is the only drug that can control glycaemia when β cells have failed and patients no longer respond to oral agents/GLP-1 analogues [269]. However, insulin therapy is associated with weight gain and hypoglycemia. In ACCORD [119] diabetic patients gained 3 kg over 3 years, and in the study by Henry et al. [270] T2DM patients gained 8.7 kg over 6 months. Weight gain results from reduction in glucosuria [271] and appetite stimulation with increased food intake. The least weight gain is observed with long-acting basal analogues (glargine and detemir) compared to short-acting insulins and NPH [230,271–274]. Weight gain associated with insulin therapy can aggravate insulin resistance, necessitating an increase in insulin dose to maintain glycemic control.

Hypoglycemia is another obstacle with insulin therapy [119]. Addition of basal insulin to poorly controlled T2DM patients increases the incidence of hypoglycemia to ~4–5 events per patient year [229,272]. If premeal rapid-acting insulin must be added to achieve glycemic control, the incidence of hypoglycemia and weight gain increases markedly [275]. Thus, in ACCORD [119] insulin therapy was associated with an annualized hypoglycemic rate requiring medical assistance of 3.1% and in the Henry study the incidence of hypoglycemia was even greater, 4.1 events per month per patient in the first month, decreasing to 1.3 events per month by 6 months [270].

There has been renewed interest in use of intensive insulin therapy to “preserve β-cell function.” In a large Swedish trial, newly diagnosed T2DM patients received premixed 30/70 insulin or glibenclamide. After 1 year, glucagon-stimulated C-peptide was significantly increased in the insulin-treated, but not glibenclamide group despite similar HbA1c reduction. After two years, HbA1c increased in the glibenclamide group and remained unchanged in the insulin group [276]. Similar results were observed in a 6-month study with basal insulin in newly diagnosed T2DM patients and improved glycemic control was attributed to enhanced β-cell function [277]. Even short courses of intensive insulin therapy can cause prolonged improvement in β-cell function in newly diagnosed T2DM patients [278]. The beneficial effect of intensive insulin therapy on β-cell function results from correction of lipotoxicity [279] and/or glucotoxicity [280].

When comparing add-on insulin to other therapeutic options, our personal belief, based upon extensive clinical experience, is that a long-acting insulin should not be added unless the subject has had a trial with a GLP-1 receptor agonist and failed.

Triple combination therapy with long-acting insulin, GLP-1 analogue, and metformin in patients with partial response to GLP-1 analogue plus metformin is very effective in achieving normoglycemia, reducing insulin dose, preventing weight gain, and reducing hypoglycemic risk [237,238,240–242].

Conclusion and recommendations

Treating diabetic patients requires in-depth knowledge of the benefits and risks of the multiple therapeutic agents, their mechanisms of action, and the patient’s needs, lifestyle, and associated medical disorders. Emphasis should be placed on medications that ameliorate insulin resistance and prevent progressive β-cell failure to achieve a durable reduction in HbA1c. T2DM is a multifactorial disease and antidiabetic medications must reverse the basic underlying pathophysiological disturbances rather than solely reduce blood glucose...
sulfonylureas are associated with progressive
UKPDS and multiple other studies have demonstrated that
they promote weight gain and cause hypoglycemia. Moreover,
than GLP-1 analogues in reducing HbA1c and do not promote
cardiovascular risk factors. DPP-4 inhibitors are less effective
failstoachievethedesiredHbA1cgoal,we recommendaddition
does not cause hypoglycemia [102,181].

Metformin is recommended by EASD/ADA/AACE as initial
therapy in newly diagnosed T2DM individuals. It reduces the
elevated rate of basal HGP, may reduce cardiovascular risk, and
does not cause hypoglycemia. However, it lacks effect on the
β cell and its effect on HbA1c is not durable [72,133–137,282].
In newly diagnosed T2DM subjects with HbA1c >8.0%, metfonrin
monotherapy will not achieve adequate glycemic control
and combination therapy should be initiated, preferably with
pioglitazone (15–30 mg day⁻¹) ± GLP-1 analogue or DPP-4i.
All of these options will achieve the desired glycemic goal
(HbA1c <6.5–7.0%) in most (70–80%) patients. Only TZDs
[161] and GLP-1 analogues [211] have been shown to produce
durable HbA1c reduction. Moreover, pioglitazone (PROactive
and US phase 3 studies) has proven cardiovascular efficacy and
does not cause hypoglycemia [102,181].

If initial combination therapy with metformin plus pioglitazone
or addition of pioglitazone to metformin-failing subjects
fails to achieve the desired HbA1c goal, we recommend addition
of a GLP-1 analogue. GLP-1 analogues effectively lower HbA1c
without hypoglycemia, promote weight loss, and improve many
cardiovascular risk factors. DPP-4 inhibitors are less effective
than GLP-1 analogues in reducing HbA1c and do not promote
weight loss but have an excellent safety profile. We do not
advocate add-on therapy with sulfonylurea or insulin because
they promote weight gain and cause hypoglycemia. Moreover,
UKPDS and multiple other studies have demonstrated that
sulfonylureas are associated with progressive β-cell failure and
loss of glycemic control.

In summary, sequential therapy with metformin, addition
of sulfonylurea with subsequent addition of insulin represents
the “treat to fail” algorithm. In contrast, initial combination ther-
apy with metformin, pioglitazone, and/or GLP-1 receptor ago-
nist provides a more rational physiologic approach and repres-
ents the approach “most likely to succeed.”

References

1 DeFronzo RA: Lilly Lecture 1987. The triumvirate: beta-cell,
muscle, liver. A collision responsible for NIDDM. Diabetes 1988;
2 DeFronzo R: Pathogenesis of type 2 diabetes: metabolic and mole-
cular implications for identifying diabetes genes. Diabetes Reviews
3 DeFronzo RA: Banting Lecture. From the Triumvirate to the Omi-
Nvous Octet: a new paradigm for the treatment of type 2 diabetes
4 Abdul-Ghani MA, Jenkinson CP, Richardson DK, et al.: Insulin
secretion and action in subjects with impaired fasting glucose and
impaired glucose tolerance: results from the Veterans Administra-
and glucose intolerance: results from the San Antonio metabolism
in subjects spanning the range from normal glucose tolerance to
7 Abdul-Ghani MA, Tripathy D, DeFronzo RA: Contributions of
beta-cell dysfunction and insulin resistance to the pathogenesis of
impaired glucose tolerance and impaired fasting glucose. Diabetes
Care 2006;29:1130–1139.
8 Groop L, Lyssenko V: Genes and type 2 diabetes mellitus. Current
9 Lyssenko V, Almgren P, Anevski D, et al.: Predictors of and longi-
tudinal changes in insulin sensitivity and secretion preceding onset
10 Weyer C, Tataranni PA, Bogardus C, Pratley RE: Insulin resistance
and insulin secretory dysfunction are independent predictors of
worsening of glucose tolerance during each stage of type 2 diabetes
11 Haffner SM, Miettinen H, Stern MP: Insulin secretion and resis-
tance in nonobese Mexican Americans and non-Hispanic whites
secretory dysfunction as precursors of non-insulin-dependent dia-
betes mellitus. Prospective studies of Pima Indians. New England
13 Dowse GK, Zimmet PZ, Collins VR: Insulin levels and the
45:1367–1372.
14 DeFronzo RA, Ferrannini E, Simonson DC: Fasting hyperglycemia
in non-insulin-dependent diabetes mellitus: contributions of
excessive hepatic glucose production and impaired tissue glucose
fatty acid metabolism in non-insulin-dependent diabetes mellitus.
Evidence for multiple sites of insulin resistance. Journal of Clinical
16 Ferrannini E, Simonson DC, Katz LD, et al.: The disposal of an
oral glucose load in patients with non-insulin-dependent diabetes.
Metabolism 1988;37:79–85.
17 Pendergrass M, Bertoldo A, Bonadonna R, et al.: Muscle glucose
transport and phosphorylation in type 2 diabetic, obese nondia-
betic, and genetically predisposed individuals. American Journal of
18 DeFronzo RA, Gunnarsson R, Bjorkman O, et al.: Effects of
insulin on peripheral and splanchnic glucose metabolism in
noninsulin-dependent (type II) diabetes mellitus. Journal of
19 DeFronzo R, Deibert D, Hendler R, et al.: Insulin sensitivity and
insulin binding to monocytes in maturity-onset diabetes. Journal of
20 James WP: The fundamental drivers of the obesity epidemic. Obes-
21 DeFronzo RA, Soman V, Sherwin RS, et al.: Insulin binding to
monocytes and insulin action in human obesity, starvation, and


62 Bays H, Mandarino L, DeFronzo RA: Role of the adipocyte, free fatty acids, and ectopic fat in pathogenesis of type 2 diabetes mellitus: peroxisomal proliferator-activated receptor agonists provide a rational therapeutic approach. JCEM 2004; 89:463 – 478.


87 DeFronzo RA, Okerson T, Viswanathan P, et al.: Effects of exenatide versus sitagliptin on postprandial glucose, insulin
and glucagon secretion, gastric emptying, and caloric intake: a randomized, cross-over study. *Current Medical Research and Opinion* 2008;24:2943–2952.


96 Garber AJ, Larsen J, Schneider SH, et al., Glyburide/Metformin Initial Therapy Study G: Simultaneous glyburide/metformin therapy is superior to component monotherapy as an initial pharmacological treatment for type 2 diabetes. *Diabetes, Obesity and Metabolism* 2002;4:201–208.


126 Hambrock A, de Oliveira Franz CB, Hiller S, Oswald H: Glibenclamide-induced apoptosis is specifically enhanced by expression of the sulfonylurea receptor isoform SUR1 but not by expression of SUR2B or the mutant SUR1(M1289T). *Journal of Pharmacology and Experimental Therapeutics* 2006;316:1031 – 1037.


175 Karamanos B, Thanopoulou A, Drossinos V, et al.: Study comparing the effect of pioglitazone in combination with either metformin
or sulphonylureas on lipid profile and glycaemic control in patients with type 2 diabetes (ECLA). *Current Medical Research and Opinion* 2011;27:303–313.


227 Gerstein HC, Yale JF, Harris SB, et al.: A randomized trial of adding insulin glargine vs. avoidance of insulin in people with Type 2 diabetes on either no oral glucose-lowering agents or submaximal doses of metformin and/or sulfonylureas. The Canadian INSIGHT (Implementing New Strategies with Insulin Glargine for Hyperglycaemia Treatment) Study. *Diabetic Medicine* 2006;23:736–742.


237 Arnolds S, Dellweg S, Clair J: Further improvement in postprandial glucose control with addition of exenatide or sitagliptin to


266 Pratley RE, Rosenstock J, Pi-Sunyer FX, et al.: Management of type 2 diabetes in treatment-naive elderly patients: benefits and


CHAPTER 47
New drugs for the treatment of diabetes mellitus

Clifford J. Bailey
School of Life and Health Sciences, Aston University, Birmingham, UK

Key points
- The heterogeneous and progressive nature of type 2 diabetes requires a variety of therapeutic options to control blood glucose.
- Longer-acting glucagon-like peptide-1 (GLP-1) receptor agonists and hybrid peptides that combine the properties of GLP-1 with other glucoregulatory peptides are being developed.
- New dipeptidyl peptidase-4 (DPP-4) inhibitors, glucokinase activators and agonists for fatty acid receptors such as GPR40 have been shown to improve endocrine functions of the pancreatic islets.
- Proof of principle has been established for nonpeptide GLP-1 receptor agonists, insulin mimetics, and sirtuin activators.
- Novel approaches to potentiate insulin action include phosphatase inhibitors and compounds that increase insulin receptor and early postreceptor signaling.
- Various adipokines and fibroblast growth factor 21 have been identified as possible targets or templates for development of new glucose-lowering therapies.
- Selective peroxisome proliferator-activated receptor (PPAR) modulators, glucagon receptor antagonists and cellular glucocorticoid inhibitors have shown potential.
- Agents that alter cellular energy metabolism (e.g. adenosine monophosphate-activated protein kinase activators) or agents that reduce glucose output (e.g. fructose 1,6-bisphosphatase inhibitors) are receiving attention.
- Sodium-glucose co-transporter-2 (SGLT2) inhibitors act independently of insulin to eliminate excess glucose via the urine.

Introduction

This chapter reviews recently approved and prospective agents for the treatment of hyperglycemia, including agents in early clinical development and preclinical compounds that indicate potential new therapeutic approaches. The main focus is towards T2DM where differently acting therapies are required to address the clinical heterogeneity that results from multivariable genetic and environmental factors and a progressive natural history [1]. Moreover, the spectrum of complications and comorbidities associated with T2DM restricts treatment options, especially at more advanced stages, and best management with currently available agents is often unable to reinstate and sustain metabolic normality and avoid or contain complications [2,3].

Approaches to new antidiabetic agents

While looking to the future, this chapter acknowledges the value of integrating new therapies into the existing framework of treatments for T2DM [3,4]. Thus lifestyle measures provide the foundation upon which drug treatments are added, noting that early interventions to address endocrine and metabolic disturbances, limit glucotoxicity and lipotoxicity, and provide comprehensive cardiovascular risk reduction are all important to achieve long-term benefits. Because diabetes is a lifetime imposition, therapies must have a good safety profile, be well tolerated and easily administered, and carry minimal risk of serious hypoglycemia [5]. Ideally, new therapies will counter the progressive decrements in metabolic control and offer novel mechanisms with additive efficacy when used in combination with other agents. Other benefits sought include weight control and pharmacokinetic properties that favor use in vulnerable groups such as the elderly and frail, or those with renal, liver, neuropathic or cardiovascular diseases.

The preclinical and clinical phases of drug development from discovery to marketing authorization are summarized in Table 47.1. The process may take 10–15 years and cost US$ 500–2000 million, with most expenditure during the phase 3 clinical trials. Additional trials may be required after marketing authorization, usually as part of a pharmacovigilance program, and further studies may be designed to acquire new indications, develop new formulations or add single-tablet fixed-dose combinations or single-injection combinations [5].

To provide structure to this review, new blood glucose-lowering agents have been organized according to their main
Table 47.1 Stages in the development of a new drug

Preclinical stages

New chemical entity (NCE)

Preclinical phases
- Identification, extraction/synthesis, chemical characterization and patenting of compounds
- Genotoxicity testing. Screening for biologic activity in vitro and in vivo in animals
- Preclinical pharmacology, mode of action, pharmacodynamics (activity, safety, tolerance), pharmacokinetics (bioavailability, distribution, metabolism, elimination) and toxicity in ≥2 mammalian species

Clinical stages

Investigational new drug (IND) application: permission to begin clinical studies

Phase 1
- First administration to a small number of healthy human volunteers. Dose ranging, vital signs, pharmacodynamics, pharmacokinetics, drug interactions, safety

Phase 2
- First trials in small numbers of patients. Dose ranging, efficacy, further pharmacodynamics, pharmacokinetics, and safety

Phase 3
- Trials in larger numbers of patients. Multicenter trials, comparative trials with other treatments, efficacy, further pharmacodynamics, and safety (including meta-analysis of cardiovascular outcomes)

New drug application (NDA): permission to market as a drug

Phase 4
- Use in medical practice. Additional trials (similar to phase 3), postmarketing surveillance, adverse events reporting, use in special subgroups (e.g. elderly and frail people)

mode or site of action. Principal actions are to enhance insulin secretion or suppress glucagon secretion, improve insulin sensitivity, independently suppress hepatic glucose output, impair counterregulatory mechanisms, delay carbohydrate digestion and intestinal glucose absorption, modulate lipid metabolism or adipose tissue activity, or directly promote glucose metabolism or increase renal glucose elimination (Figure 47.1). For comparison, and to indicate opportunities for complementary and additive use, the main actions of existing agents are summarized in Table 47.2.

Inhibitors of intestinal carbohydrate digestion and absorption

Fiber-rich diets and fiber supplements slow the digestion and absorption of dietary carbohydrate, reducing postprandial hyperglycemia while often reducing insulin concentrations. Soluble fibers such as gums, pectins, and hemicelluloses appear to be more effective than insoluble fibers such as celluloses and wheat bran, providing a greater barrier for diffusion of digestive enzymes and liberated saccharides. However, soluble fibers such as the galactomannan guar gum (E412) and fruit pectins are already common food thickeners, and normal balanced diets with fruits and vegetables should contain plenty of fiber. The effect of further fiber is generally modest, although glucomannan soluble fibers supplements are widely used. It is noted that fiber-rich diets can be used in conjunction with any antidiabetes drug therapy to reduce prandial hyperglycemic excursions, and they may usefully reduce interprandial hypoglycemia in insulin-treated patients [6].

Slowing carbohydrate digestion with reversible competitive inhibitors of α-glucosidase enzymes, such as acarbose, miglitol and voglibose, reduces postprandial hyperglycemia and often reduces insulin concentrations in individuals who consume a diet rich in complex carbohydrate (Figure 47.2). Inhibitors of α-amylase enzymes have been considered as agents to slow the hydrolysis of dietary starch, but the effects have been too unpredictable for routine therapeutic use, running the risk that undigested sugars might enter the large bowel and undergo fermentation [6].

Initiators of insulin secretion

The progressive decline of β-cell function in T2DM involves loss of the acute first-phase insulin secretory response to glucose, impaired processing of proinsulin to insulin, an abnormal pulsatile rhythm of basal insulin secretion, and an extended second phase of glucose-induced insulin secretion that diminishes in magnitude as the extent of postprandial hyperglycemia increases (see Chapters 24 and 25). In advanced stages of T2DM there are also reductions in β-cell mass and insulin biosynthesis [1]. Insulin secretagogues can be categorized into initiators (e.g. sulfonylureas), which stimulate insulin secretion on their own or require only low glucose concentrations, and
potentiators (e.g. GLP-1 receptor agonists), which enhance the effect of raised concentrations of glucose and other initiating secretagogues but do not initiate insulin release on their own.

\( \text{K}_{\text{ATP}} \) channel closers

The biochemical route map within the islet \( \beta \) cell through which glucose stimulates insulin secretion is illustrated in Figure 47.3. Sulfonylureas and meglitinides activate this route part way along by binding to the so-called sulfonylurea receptor-1 (SUR1) in the plasma membrane. SUR1 is part of the adenosine triphosphate (ATP)-sensitive potassium channel (\( \text{K}_{\text{ATP}} \) channel) comprising an octameric complex of four Kir6.2 pore-forming complexes (inwardly-rectifying potassium channels) surrounded by four SUR1 molecules. Although sulfonylureas and meglitinides bind at different sites on SUR1 the effect is much the same as binding of ATP to the nucleotide-binding domains, namely to close the Kir6.2 pore [7]. This prevents \( \text{K}^+ \) efflux, causing localized depolarization of the plasma membrane which opens voltage-gated (L-type) calcium channels. The ensuing influx of \( \text{Ca}^{2+} \) ions increases the cytosolic calcium ion concentration which activates calcium-sensitive proteins that trigger the exocytosis of insulin-containing secretory granules. SUR1–Kir6.2 channels are also located within the membranes of mitochondria and possibly other organelles, suggesting that sulfonylureas and meglitinides could act at these sites to influence insulin release. Since this insulin secretory mechanism can operate at low glucose concentrations it carries a risk of interprandial hypoglycemia, whereas an ideal insulin secretagogue would restore \( \beta \)-cell sensitivity to raised glucose and support adequate biosynthesis, processing and secretion of insulin without stimulating insulin secretion at low glucose concentrations.

Several different types of compounds that close \( \text{K}_{\text{ATP}} \) channels have been considered as potential new insulin secretagogues. Only the meglitinide derivative mitiglinide (KAD-1229) has proceeded through clinical development, receiving approval for routine use in some countries [6]. Other compounds that close \( \text{K}_{\text{ATP}} \) channels such as the morpholinoguanidine BTS67582 and certain imidazolines have not proceeded beyond early clinical studies, although several of the imidazolines also bind to I1, I2 and probably other receptors, which may be involved in insulin secretion. There are also imidazolines such as BL11282 that stimulate glucose-induced (but not basal) insulin secretion independently of \( \text{K}_{\text{ATP}} \) channels, probably via effects on protein kinases. Some nonselective alpha2-adrenergic receptor antagonists such as phentolamine, which can reduce the tonic suppression of insulin secretion mediated through alpha2-adrenergic activation, may also act to close \( \text{K}_{\text{ATP}} \) channels [6].

Although closure of \( \text{K}_{\text{ATP}} \) channels initiates insulin secretion it does not increase proinsulin biosynthesis. Since increased nutrient metabolism within the \( \beta \) cell will increase proinsulin biosynthesis and close \( \text{K}_{\text{ATP}} \) channels via increased ATP production, the stimulation of mitochondrial nutrient metabolism has been evaluated as a potential route to increase insulin
Table 47.2 Existing agents with approved indications as blood glucose-lowering agents

<table>
<thead>
<tr>
<th>Class and examples</th>
<th>Main mechanisms of action</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Oral</strong></td>
<td></td>
</tr>
<tr>
<td>Biguanide</td>
<td>Reduce hepatic glucose production, improve insulin action, increase gut-liver glucose cycling and other insulin-independent effects</td>
</tr>
<tr>
<td>Metformin</td>
<td></td>
</tr>
<tr>
<td>Sulfonylureas</td>
<td>Stimulate insulin secretion by binding to SUR-1 on β cells and closing K$_{ATP}$ channels (effect typically lasts 6–24 h)</td>
</tr>
<tr>
<td>Glibenclamide (Glyburide)</td>
<td></td>
</tr>
<tr>
<td>Gliclazide</td>
<td></td>
</tr>
<tr>
<td>Glimepiride</td>
<td></td>
</tr>
<tr>
<td>Glipizide</td>
<td></td>
</tr>
<tr>
<td>Tolbutamide</td>
<td></td>
</tr>
<tr>
<td>Meglitinides (Prandial insulin releasers, meglitinides)</td>
<td>Stimulate insulin secretion by binding to SUR-1 on β cells at a different site to sulfonylureas and closing K$_{ATP}$ channels (rapid effect, typically lasts &lt;6 h)</td>
</tr>
<tr>
<td>Miglitolinide</td>
<td></td>
</tr>
<tr>
<td>Nateglinide</td>
<td></td>
</tr>
<tr>
<td>Repaglinide</td>
<td></td>
</tr>
<tr>
<td>DPP-4 inhibitors (Glipitins)</td>
<td>Inhibit enzyme dipeptidyl peptidase-4 prolonging circulating half-lives of some incretin hormones such as GLP-1 which enhance prandial insulin secretion and exert other glucoregulatory effects</td>
</tr>
<tr>
<td>Alogliptin</td>
<td></td>
</tr>
<tr>
<td>Linagliptin</td>
<td></td>
</tr>
<tr>
<td>Saxagliptin</td>
<td></td>
</tr>
<tr>
<td>Sitagliptin</td>
<td></td>
</tr>
<tr>
<td>Vildagliptin</td>
<td></td>
</tr>
<tr>
<td>Thiazolidinediones (Glitazones. TZDs)</td>
<td>Improve insulin sensitivity mainly via activation of peroxisome proliferator-activated receptor-γ (PPARγ)</td>
</tr>
<tr>
<td>Pioglitazone</td>
<td></td>
</tr>
<tr>
<td>Rosiglitazone</td>
<td></td>
</tr>
<tr>
<td>SGLT2 inhibitors</td>
<td>Inhibit mainly renal sodium-glucose co-transporter 2 to enable increased glucose elimination in the urine</td>
</tr>
<tr>
<td>Canagliflozin</td>
<td></td>
</tr>
<tr>
<td>Dapagliflozin</td>
<td></td>
</tr>
<tr>
<td>Empagliflozin</td>
<td></td>
</tr>
<tr>
<td>Alpha-glucosidase inhibitors</td>
<td>Slow rate of carbohydrate digestion</td>
</tr>
<tr>
<td>Acarbose</td>
<td></td>
</tr>
<tr>
<td>Miglitol</td>
<td></td>
</tr>
<tr>
<td>Voglibose</td>
<td></td>
</tr>
<tr>
<td>Dopamine agonist</td>
<td>Improve circadian rhythmicity of glycemic control with suppression of hepatic glucose production</td>
</tr>
<tr>
<td>Bromocriptine</td>
<td></td>
</tr>
<tr>
<td>Bile sequestrant</td>
<td>Possibly alter secretion of incretin hormones such as GLP-1</td>
</tr>
<tr>
<td>Colesevelam</td>
<td></td>
</tr>
<tr>
<td><strong>Subcutaneous injection</strong></td>
<td>Activate GLP-1 receptors, mimic action of GLP-1 to enhance prandial insulin secretion, reduce prandial glucagon secretion, delay gastric emptying and exert satiety effect</td>
</tr>
<tr>
<td>GLP-1 receptor agonists</td>
<td></td>
</tr>
<tr>
<td>Exenatide</td>
<td>Slow gastric emptying, suppress glucagon secretion, possible satiety effect</td>
</tr>
<tr>
<td>Liraglutide</td>
<td></td>
</tr>
<tr>
<td>Lixisenatide</td>
<td></td>
</tr>
<tr>
<td>Amylin analogue</td>
<td></td>
</tr>
<tr>
<td>Pramlintide</td>
<td></td>
</tr>
</tbody>
</table>

Table 47.2 (continued)

<table>
<thead>
<tr>
<th>Class and examples</th>
<th>Main mechanisms of action</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Insulin</strong></td>
<td>Reduce hepatic glucose output</td>
</tr>
<tr>
<td>Rapid-acting:</td>
<td>Increase peripheral glucose uptake</td>
</tr>
<tr>
<td>Aspart, Glulisine, Lispro</td>
<td>Increase glucose metabolism</td>
</tr>
<tr>
<td>Short-acting:</td>
<td>Decrease lipolysis</td>
</tr>
<tr>
<td>Actrapid, Humulin S, Insuman Rapid</td>
<td>Increase lipogenesis, Favor protein anabolism</td>
</tr>
<tr>
<td><strong>Intermediate</strong></td>
<td></td>
</tr>
<tr>
<td>Insulatard, Humulin I</td>
<td></td>
</tr>
<tr>
<td>Long-acting:</td>
<td></td>
</tr>
<tr>
<td>Degludec, Detemir, Glargine</td>
<td></td>
</tr>
<tr>
<td>Biphasic (pre-mixed): Humalog, Humulin, M3, Novomix</td>
<td></td>
</tr>
<tr>
<td><strong>DPP-4, dipeptidyl peptidase-4; GLP-1, glucagon-like peptide-1; K$_{ATP}$ channels, adenosine triphosphate-sensitive potassium channels; PPARγ, peroxisome proliferator-activated receptor-gamma; SGLT, sodium-glucose co-transporter; SUR-1, sulfonylurea receptor-1; TZDs, thiazolidinediones.</strong></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 47.2** Alpha-glucosidase inhibitors slow carbohydrate digestion by competitive inhibition of α-glucosidase enzymes at the luminal surface of brush border membranes of enterocytes in the small intestine. Alpha-glucosidase inhibitors in present use are acarbose, miglitol, and voglibose. These agents have different affinities for specific α-glucosidases: the binding affinity of acarbose competitively inhibits glucoamylase > sucrase > dextrinase; miglitol and voglibose potentially inhibit sucrase, while voglibose shows more potent inhibition of other α-glucosidases than acarbose. Acarbose also weakly inhibits alpha-amylose.

Secretion alongside proinsulin biosynthesis. Succinate esters, for example, fulfil this role, but their low enteral bioavailability and short duration of action have precluded clinical development, and their propensity to fuel hepatic gluconeogenesis is a further deterrent. Stimulation of mitochondrial succinyl-coenzyme A (CoA) synthetase to generate ATP and guanosine triphosphate (GTP) has also been considered, but has proved difficult to specifically target the β cell [6].
Glucokinase activators

The enzyme glucokinase (GK, EC 2.7.1.1) is expressed by islet \(\beta\) cells and hepatocytes where it phosphorylates glucose to glucose-6-phosphate. Since these cells can take up glucose (via the glucose transporters GLUT2, and GLUT1 in human islets) approximately in proportion to the circulating glucose concentration, GK constitutes the rate-limiting step for metabolism of glucose in these cells [8]. Thus GK activators have been considered as a therapeutic approach to enhance glucose metabolism in both cell types. This would enhance glucose-induced insulin secretion and proinsulin biosynthesis through ATP production by the \(\beta\) cells, and facilitate glucose utilization or glycogen storage in liver. Islet \(\alpha\) cells also express GK which appears to assist the glucose-dependent suppression of glucagon secretion.

Several specific small molecule allosteric activators of GK (e.g. various acetamides and benzamides) have been shown to increase insulin secretion and improve glucose homeostasis in animal models of non-insulin-dependent diabetes [8]. Although this effect is actually potentiating the action of glucose, it can operate at low glucose concentrations, so the effect is similar to an initiator, and if this effect is not rapidly terminated when the glucose concentration subsides, there is a risk of continued insulin secretion and hypoglycemia.

Glucokinase is regulated slightly differently in the liver compared with islet \(\beta\) cells. In the liver GK becomes activated by the influx of glucose which slowly dissociates the enzyme from a regulatory binding protein, whereas islet cells do not express the binding protein [8]. Thus the action of GK in liver can be enhanced and prolonged by allosteric activators and by molecules that prevent association of GK with or cause dissociation of GK from its regulatory binding protein. Several such agents have been assessed in clinical trials but the tight therapeutic index to avoid hypoglycemia and the lack of durable efficacy have so far prevented completion of a development program. Liver-selective GK activators are being considered as a way of avoiding hypoglycemia, but the potential risk of excess liver glycogen and hepatic lipogenesis remains to be determined.

Potentiators of insulin secretion

Agents that potentiating nutrient-induced insulin secretion should act mainly to reduce postprandial hyperglycemia with less risk of interprandial hypoglycemia than agents that can initiate insulin secretion at low glucose concentrations.

Incretins

Incretin hormones are released from the gut during feeding and enhance nutrient-induced insulin secretion and suppress glucagon secretion. The main incretins are glucagon-like peptide-1 (GLP-1) and gastric inhibitory polypeptide (GIP; also known as glucose-dependent insulinotropic peptide). They appear to alter islet cell function partly via neural pathways activated during their passage through the portal circulation.
Table 47.3 Glucoregulatory effects of glucagon-like peptide-1 (GLP-1) and gastric inhibitory polypeptide (GIP; also known as glucose-dependent insulinotropic peptide)

<table>
<thead>
<tr>
<th>Effect</th>
<th>GLP-1</th>
<th>GIP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Released in response to a mixed meal</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Lower blood glucose</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Glucose-dependent stimulation of insulin secretion</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Suppress glucagon secretion</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Extrapancreatic glucose-lowering actions</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Extend β-cell mass and survival</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Suppress gastric acid secretion</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Inhibition of hepatic insulin extraction</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Enhance satiety</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Reduce body weight</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

and partly via direct effects on the islets. G-protein coupled receptors (GPCRs) in the β-cell membrane, when activated by their incretin ligands, potentiate mainly glucose-induced insulin secretion and proinsulin biosynthesis through activation of adenylate cyclase, increasing production of cyclic adenosine monophosphate (cAMP) and protein kinase A (PKA) [9]. Effects on Ca\(^{2+}\) ion distribution within the islet β cells may also be involved. Both GLP-1 and GIP are reported to promote β-cell mass in animal models, possibly slowing β-cell apoptosis and increasing β-cell neogenesis by increased expression of the transcription factor pancreatic duodenal homeobox-1 (PDX-1), which promotes proliferation and differentiation of ductal progenitor cells. However, these latter effects on β-cell mass have not been confirmed in human diabetes.

Significantly impaired secretion of GLP-1 or GIP has not been consistently observed in T2DM, but the potentiation of glucose-induced insulin secretion in T2DM with pharmacologic doses of GLP-1 remains stronger than with GIP, which accounts in part for the therapeutic use of GLP-1 rather than GIP. Other properties of GLP-1 are very important: GLP-1 suppresses glucagon secretion in a glucose-dependent manner, slows gastric emptying, and exerts a satiety effect that can facilitate weight loss (Table 47.3). However, rapid degradation of GLP-1, mostly by the enzyme dipeptidyl peptidase-4 (DPP-4) and the renal clearance of GLP-1 (plasma half-life <2 min) limits use of the peptide itself as a form of therapy [9,10].

**GLP-1 receptor agonists**

DPP-4 cleaves the N-terminal dipeptide if there is an Ala residue (as in GLP-1) or Pro residue at the N2 position. Of the currently available GLP-1 receptor agonists, exenatide (exendin-4) and lixisenatide are GLP-1 analogues with a Gly residue at the N2 position and substitutions at other positions (Figure 47.4), making these molecules resistant to degradation by DPP-4 and endopeptidases, thereby extending the circulating half-life to several hours. Liraglutide is GLP-1 (7-37) with Lys34 replaced by Arg34, and Lys26 attached via a glutamate residue to a C16 hexadecanoyl (palmitoyl) fatty acid chain. The fatty acid chain attaches to albumin, protecting against degradation by DPP-4 [10,11].

Each of these molecules retains a strong affinity for the GLP-1 receptor and is delivered by subcutaneous injection. Exenatide is given twice daily before the main meals or as a once-weekly depot formulation in which the molecule is embedded within polylactide-glycolide microspheres. Lixisenatide is given once or twice daily and liraglutide once daily. The glucose-dependent nature of the insulin-releasing and glucagon-suppressing effects is reflected in the limited risk of hypoglycemia, while the satiety effect has assisted weight loss and favors use in obese patients. Nausea related to the delayed gastric emptying is not uncommon as a temporary early side effect. Several further once-weekly and potentially longer lasting injectable GLP-1 analogues are advanced in clinical trials: for example albiglutide is two GLP-1 analogues linked to albumin, while dulaglutide is two GLP-1 analogues linked to IgG. Depot formulations releasing GLP-1 analogues over one or several months have also entered clinical trials, and noninjected routes of administration of GLP-1 analogues including transdermal, buccal and inhaled are being explored [11,12].

**GIP analogues**

Although GIP potentiates nutrient-induced insulin secretion it also increases glucagon release and promotes lipid deposition. Several DPP-4-resistant GIP analogues have been shown to modestly improve glucose control without weight...
Combinatorial, chimeric, and hybrid peptides

Given the potential for combination of GLP-1 with either a GIP agonist or antagonist, combinatorial (linked) native or analogue versions of these peptides have been evaluated. In general these have shown little additional benefit beyond that of a GLP-1 receptor agonist alone, but the concept of combining two or more peptides into a single molecule is now receiving wider therapeutic attention [14].

Like GLP-1, some other incretins show homology or commonality through precursor biosynthesis with glucagon. While glucagon is well known to increase hepatic glucose production (and the potential of glucagon receptor antagonists is considered later), glucagon also assists weight loss through reduced meal size, increased lipolysis, and increased energy expenditure. Oxyntomodulin, derived from preproglucagon, is a weak agonist of receptors for GLP-1 and glucagon, suppressing food intake, increasing energy expenditure, and illustrating a naturally occurring form of combination. For pharmacologic extension of this overlap, chimeric peptides (part GLP-1 agonist, part glucagon agonist) and hybrid peptides (design based on selected partial sequences of GLP-1 and glucagon) have been constructed with strong GLP-1 receptor agonism and weaker glucagon agonism. These have shown potentiation of glucose-induced insulin secretion and weight loss while overriding the hyperglycemic effect of glucagon. Duration of potency of the molecules has been prolonged by avoiding an N-terminal sequence susceptible to DPP-4 and by conjugation to polyethylene glycol. Since some effects of glucagon may be facilitated by fibroblast growth factor 21, this has also been considered as an addition [14].

Nonpeptide incretin mimetics

The clinical application of large peptides implies parenteral (probably subcutaneous) administration, and peptides raise questions regarding long-term antibody production, attenuated responses to native hormones and receptor downregulation. To address some of these issues development of nonpeptide small molecule incretin receptor agonists and antagonists is being considered. For example, nonpeptide GLP-1 receptor agonists bind to GLP-1 receptors on islet β cells and enhance glucose-dependent insulin secretion in animals, while a GIP antagonist has been shown to counter obesity and insulin resistance in animals [13,15].

**DPP-4 inhibitors**

Agents that inhibit the enzyme DPP-4 (EC 3.4.14.5) prevent rapid inactivation of GLP-1 and GIP, raising their circulating concentrations and thereby enhancing their effects [10,16]. Current DPP-4 inhibitors (“gliptins”) include sitagliptin, vildagliptin, saxagliptin, linagliptin, and alogliptin (see Chapter 48). Although there are subtle pharmacokinetic differences, therapeutic use of each of these agents will produce almost complete inhibition of DPP-4 activity for most of a 24-h period, and they each improve glycemic control without weight gain. Their effects are attributed largely to increased concentrations of GLP-1 which are raised about two- to threefold, although it is noted that DPP-4 inhibition may also raise GIP concentrations. For comparison, injected GLP-1 receptor agonists generate circulating values equivalent to about 10-fold higher than endogenous GLP-1. These differences could explain why DPP-4 inhibitors do not consistently produce the magnitude of effects of injected GLP-1 receptor agonists and do not cause significant weight reductions. However, DPP-4 inhibitors do not cause initial nausea through delayed gastric emptying.

DPP-4 exists unattached in the circulation and tethered to cell membranes especially of endothelia in capillaries of the gastrointestinal tract. The protease activity of DPP-4 degrades a range of biologically active peptides in addition to incretins, where there is an N2 Ala or Pro. Susceptible peptides include substance P, bradykinin, peptide YY, neuropeptide Y, pituitary adenylate cyclase-activating peptide, insulin-like growth factor-1, and various interleukins and monocyte chemo-attractant proteins [16,17]. Despite affecting this wide range of peptides, to date DPP-4 inhibitors have not shown significant adverse effects during substantial clinical use. Also, DPP-4 is the lymphocyte cell surface protein CD26 required for the co-stimulation response to recall antigens, but its immunologic role does not appear to be interrupted by small molecule inhibitors of its peptidase activity. This lack of interference with other biologic functions has prompted development of long-acting (e.g. once-weekly) DPP-4 inhibitors such as MK-3102 [18].

**Fatty acid receptor agonists**

Pancreatic β cells take up fatty acids which serve as metabolic substrates: acutely they enhance insulin secretion, but chronic exposure to raised fatty acid concentrations within the cell causes detrimental lipotoxic effects. In addition to these effects certain fatty acids serve as ligands for GPCRs in the β-cell membrane. Binding to these receptors does not give a fatty acid access to the cell, but activates the GPCRs to generate intracellular signals that influence insulin secretion, mainly by actions that potentiate insulin secretion that is initiated by other factors [19]. The principal fatty acid-activated GPCRs of the β cells are GPR40 (FFAR1) and GPR119, and synthetic ligands
have been developed to activate the receptors and potentiate glucose-induced insulin secretion.

Ligand binding to GPR40 activates a subunit of the receptor that signals via phospholipase C (PLC) to enhance insulin secretion by raising intracellular Ca^{2+} partly by mobilizing Ca^{2+} from endoplasmic reticulum and partly by increasing Ca^{2+} influx across the cell membrane (Figure 47.2). Additionally the PLC increases diacylglycerol which acts via isofoms of protein kinase C to increase distal steps in the exocytotic pathway. Ligand binding to GPR119 activates a subunit of the receptor that signals via adenylate cyclase to enhance cyclic adenosine monophosphate and protein kinase A as described earlier for inrectins. GPR40 and 119 are expressed on intestinal L cells that secrete GLP-1, enabling synthetic ligands to further enhance insulin secretion via increased GLP-1. However, GPR40 and especially GPR119 are expressed by islet α cells, where they could mediate increased glucagon secretion [19,20].

**Other insulin secretion potentiators**

Tetrahydrotriazine-containing compounds, notably imeglimin, enhance glucose-induced insulin secretion and may reduce β-cell apoptosis. Additionally imeglimin acts on liver to reduce gluconeogenesis and on skeletal muscle to promote glucose uptake. Several phosphodiesterases, particularly isoform PDE-3B, are expressed by islet β cells, degrading cAMP and reducing insulin release. Selective and transient inhibition of this isoform is a possible intervention but difficult to target specially to β cells. Antagonism of alpha2-adrenoceptors remains another possibility [6].

**Increasing insulin action**

Impaired insulin action (insulin resistance) is an early and sustained feature of most presentations of adult-onset T2DM. Hence new interventions are sought to improve insulin receptor and postreceptor signaling, or to act independently of insulin to address disturbances of substrate transporters, metabolic enzymes, and other abnormalities that give rise to glucose and lipotoxicity [21]. It is noted that although some of the metabolic consequences of insulin resistance can be relieved by interventions that are independent of insulin, a basal level of insulin action is required for the genomic effects that determine expression levels for many of the cellular components that directly and indirectly control metabolic homeostasis. Excess proinsulin and insulin-like growth factor-1 represent prereceptor factors contributing to insulin resistance because they bind weakly to the insulin receptor and generate a much lesser response than insulin. Defects of insulin receptor structure are uncommon, and reductions in insulin receptor numbers are not usually rate limiting, suggesting that interventions are required to enhance insulin receptor signaling and early postreceptor signaling defects in order to increase the spectrum of insulin effects. However, rate-limiting defects can occur at more distal locations and negative feedbacks can operate that limit the full transmission of improvements at early steps in the signaling pathways. Also, most of the postreceptor pathways are not specific to insulin and impact activities as disparate as cell differentiation and apoptosis. Thus many potential sites are being explored to enhance insulin action (Figure 47.5), bearing in mind that they should circumvent rate-limiting bottlenecks without distorting the overall equanimity of signaling balance across many cellular functions [22].

**Initiating insulin receptor activity**

Therapeutic activation of the insulin receptor is presently achieved with insulin, insulin analogues and related peptides, but a convenient advance would be the replacement of these agents with small nonpeptide compounds that could be administered orally. The fungal metabolite demethylasterriquinone provided initial evidence for a nonpeptide insulin mimetic. Low concentrations (3–6 μM) of demethylasterriquinone initiated phosphorylation and tyrosine kinase activity of the B-subunit of the human insulin receptor expressed in Chinese hamster ovary (CHO) cells [23]. This in turn induced tyrosine phosphorylation and activation of insulin receptor substrate-1 (IRS1), with downstream activation of phosphatidylinositol 3-kinase (PI3K) and Akt (protein kinase B). Studies with mutated A-subunits of the insulin receptor showed that demethylasterriquinone interacted selectively with the B-subunit (without requiring insulin to bind to the A-subunit), and its activity could not be attributed to inhibition of protein tyrosine phosphatases. Demethylasterriquinone produced a range of insulin-like effects in normal tissues including increased glucose uptake by isolated rodent adipocytes and skeletal muscle, and orally administered doses of 5–25 mg·kg\(^{-1}\)·d\(^{-1}\) lowered blood glucose in insulin-resistant obese diabetic db/db mice. Various features of demethylasterriquinone are not suited to clinical development but this molecule has provided evidence for the concept for an orally active small molecule insulin mimetic. Although the low affinity binding of proinsulin and insulin-like growth factor-1 (IGF-1) to the insulin receptors can impede the effectiveness of native insulin, therapeutic advantage has been taken of cross-talk between the IGF-1 receptor and early postreceptor components of the insulin-signaling cascades to treat some rare cases of severe insulin resistance caused by genetic defects of the insulin receptor. However, extra amounts of IGF-1 carry the caution of potential proliferative and other effects, although overly raised peak concentrations of free circulating IGF-1 after subcutaneous administration have been limited by concomitant administration of the binding protein IGFBP-3 [6].

Small variations in the metabolic responses to different insulin-related peptides suggest that subtle differences in binding at the receptor A-subunit can modulate the profile of signaling activities of the B-subunit [24]. Selective insulin receptor modulation might be used to avoid potential mitogenic effects associated with activation of IGF-1 receptors. For example a monoclonal antibody has been developed that binds...
Figure 47.5 Pathways of intracellular insulin signaling showing some of the potential sites for therapeutic intervention. Source: Based on Bailey 2007 [22]. Akt, protein kinase B (PKB); AMPK, adenosine monophosphate-activated protein kinase; DAG, diacylglycerol; eNOS, endothelial nitric oxide synthase; FAs, fatty acids; FOXO1, forkhead box protein O1A; GLUT, glucose transporter isoform; Grb, growth factor receptor binding protein; GSK3, glycogen synthase kinase 3; IKKB, inhibitor kappa-B kinase beta; IL6, interleukin 6; IRS, insulin receptor substrate; JNK, c-Jun N-terminal kinase; MAPK, mitogen-activated protein kinase; mTOR, mammalian target of rapamycin; PC-1/NNP1, glycoprotein-1; PDK, phosphoinositide-dependent protein kinase; PGC-1α, PPAR co-activator-1alpha; PIP2, phosphatidylinositol-3,4-bisphosphate; PIP3, phosphatidylinositol-3,4,5-trisphosphate; PKC, protein kinase C; PPAR, peroxisome proliferator-activated receptor; PPS2C, pyruvate dehydrogenase phosphatase (protein phosphatase 2C); PTEN, protein phosphatase PTEN; PTP-1B, protein tyrosine phosphatase-1B; Raf, a serine-threonine protein kinase; Ras, a guanosine triphosphatase; ROS, reactive oxygen species; RXR, retinoid X receptor; SHIP-2, src homology-2-inositol phosphatase; SOCS-3, suppressor of cytokine signaling-3; SOS, sons of sevenless; STAT, signal transducer and activator of transcription; TNF-α, tumor necrosis factor alpha; ↑, increase.
specifically to the insulin receptor and produces predominantly metabolic effects of insulin [25].

**Potentiating insulin receptor activity**

Binding of insulin to the A-subunit of the insulin receptor produces conformational changes that enable tyrosine phosphorylation of the B-subunit. Therapeutic agents that interact with the B-subunit to augment or prolong its tyrosine phosphorylation have therefore been sought, and nonpeptide molecules that achieve this have been shown to enhance insulin action, such as TLK16998 [26].

Co-administration of insulin C-peptide together with insulin augments the metabolic effects of insulin, and there is evidence from studies in isolated liver and muscle cells that the C-peptide may enhance insulin receptor tyrosine kinase activity and phosphorylation of IRS1 and other postreceptor signaling intermediates [27].

**Protein tyrosine phosphatase-1B inhibitors**

Several protein tyrosine phosphatases (PTPs), especially PTP-1B, dephosphorylate the B-subunit of the insulin receptor and so terminate its tyrosine kinase activity [28]. These phosphatases also dephosphorylate and deactivate the insulin receptor substrates IRS1 and IRS2, whereas PTP-1B knockout mice and antisense oligonucleotides against PTP-1B increase insulin sensitivity. As proof of concept, several selective small molecule inhibitors of PTP-1B have improved glycemic control in insulin-resistant diabetic animals [29]. Inhibiting PTP-1B may also assist weight control by increasing the satiety effect of leptin receptor signaling, and improve endothelial function by improving insulin-induced endothelial nitric oxide synthase (eNOS) production [6]. Vanadium salts (reviewed later) inhibit PTP-1B, and have been shown to enhance the actions of insulin and possibly leptin.

**Inhibition of negative feedback signals**

Early components of postreceptor signaling pathways of insulin action involve phosphatidinositol-3,4,5 trisphosphate (PIP3) and phosphoinositide-dependent kinases PDK-1/2. These signaling intermediates activate certain isoforms of protein kinase C (PKC) such as PKC-theta which exerts a negative feedback by phosphorylation of a serine on the insulin receptor B-subunit and on IRS-1, preventing tyrosine phosphorylation and reducing kinase activity. Thus, inhibition of specific isoforms of PKC provides a potential means to improve insulin action, but developing highly selective inhibitors has proved difficult [6,22].

Other serine kinases that interfere with the tyrosine kinase activity of the insulin receptor B-subunit and IRS proteins are potential therapeutic targets. These include inhibitor kappa-B kinase-beta (IKKβ) and c-Jun N-terminal kinase (JNK) which contribute to the insulin resistance produced by the cytokine tumor necrosis factor-alpha. Agents that suppress these serine kinases (e.g. salicylates inhibit IKKβ) can partly relieve insulin resistance. The insulin signaling intermediate Akt activates the mammalian target of rapamycin (mTOR) which exerts a negative feedback on IRS proteins through serine phosphorylation, and the membrane ectoenzyme glycoprotein-1 (PC-1/NNP1) binds to the insulin receptor and prevents the conformational changes required for receptor autophosphorylation [6,22]. These interactions present potential sites for therapeutic intervention, though it is difficult to envisage selective interference that would not impinge on other regulatory pathways.

**Potentiation of postreceptor signals**

The phosphorylation of phosphatidylinositol 4,5-bisphosphate (PIP2) to the 3,4,5-trisphosphate (PIP3) by phosphatidylinositol 3-kinase (PI3K) is often reduced in insulin-resistant states, and increased availability of precursor inositol substrates may improve this step. Thus, D-chiro-inositol (INS-1) and the 3-methoxy analogue of D-chiro-inositol (pinitol) improve muscle glucose uptake and reduce hyperglycemia in diabetic animal models and type 2 diabetes patients [22]. D-chiro-inositol-galactosamine (INS-2) also increased insulin action in diabetic rats, possibly involving activation of pyruvate dehydrogenase phosphatase (protein phosphatase 2C; PP2C).

Another approach could be to prevent the dephosphorylation of PIP3. The phosphatase PTEN which dephosphorylates PIP3 can be inhibited by an antisense oligonucleotide which increases signaling downstream of PIP3, and improves insulin sensitivity and glycemic control in insulin-resistant animal models [30]. PIP3 is also dephosphorylated by the phosphatase SHIP-2, and partial disruption of the SHIP-2 gene in mice improves insulin sensitivity. However, interference with PIP3 dephosphorylation may disturb pathways that could carry a risk of tumor formation.

**Adipokines**

Several of the autocrine, paracrine, and endocrine factors produced by adipose tissue affect insulin action and have been considered as targets or templates for therapeutic agents (Figure 47.6) [31]. For example, adipose tissue is a source of pro-inflammatory cytokines such as TNF-alpha and interleukin-6 (IL6), which impair insulin receptor B-subunit tyrosine kinase activity through JNK, IKKB, and the suppressor of cytokine signalling-3 (SOCS-3) (Figure 47.5). However, these cytokines exert other effects that may influence glucose homeostasis, and antibodies to IL6 and TNF-alpha have shown limited efficacy as treatments for diabetes and may increase susceptibility to infection.

**Adipocyte hormones**

Of the many adipocyte hormones, leptin and adiponectin have received most attention as possible diabetes therapies.

**Leptin**

The centrally mediated satiety and thermogenic effects of leptin which facilitate weight loss, and direct peripheral effects of
leptin to improve insulin action and possibly suppress glucagon, have made this peptide and its analogues and fragments potential treatments for obesity and diabetes. However, the development of leptin resistance and leptin antibodies has so far compromised long-term efficacy [32].

Adiponectin (Acrp30)  
Because adiponectin increases insulin sensitivity, improves vascular reactivity and exerts anti-inflammatory activity, it appeals as a therapeutic template. Its presence in large amounts and with different circulating isoforms makes use of the native peptide difficult, but fragments and nonpeptide receptor agonists are under investigation [33,34].

Resistin (FIZZ3)  
This adipocyte and macrophage hormone reduces insulin sensitivity, increases pro-inflammatory cytokines and exerts several adverse vascular effects. Since immunoneutralization of resistin has improved insulin sensitivity in rodents, inhibition of resistin action is being considered as a therapeutic approach [35].

Retinol-binding protein 4 (RBP4)  
Several binding proteins that transport plasma retinoids are produced by adipocytes and liver. Increased RBP4 has been noted in the early stages of insulin resistance and obese states, and RBP4 gene knockout increases insulin sensitivity in animals, suggesting that a reduction of RBP4 might be considered to reduce the initial development of insulin resistance [36].

Visfatin, vaspin, omentin, and apelin  
Each of these adipocyte peptides has recently been reported to improve insulin action, but their therapeutic potential is not clear. Visfatin derives mainly from visceral adipocytes, and appears to activate insulin receptors, increase insulin secretion, and alter expression of several genes linked to diabetes, but clinical studies have not shown a correlation between plasma visfatin and insulin sensitivity. Vaspin (visceral adipose tissue-derived serpin), omentin, and apelin have been reported to increase insulin-stimulated glucose disposal in obese insulin-resistant rodents and isolated cells [31,37].

Fibroblast growth factor 21  
Fibroblast growth factor 21 (FGF21), a peptide secreted mostly from liver, muscle and adipose tissue, plays an important role during starvation by promoting hepatic gluconeogenesis, fatty acid oxidation, and ketogenesis. Pharmacologic doses of FGF21 administered to obese diabetic animals enhance insulin sensitivity, reduce hyperglycemia (despite increased gluconeogenesis), promote energy expenditure and weight loss, and improve the blood lipid profile without any apparent mitogenic activity. Although plasma FGF21 concentrations may be raised in human obesity and T2DM, indicating possible FGF21 resistance, clinical studies have noted that administration of FGF21 analogues can improve the lipid profile and produce a modest glucose-lowering effect [38]. Recent evidence indicates that FGF21 may improve insulin action via increased adiponectin secretion, and it is noted that FGF21 and adiponectin stimulate PPARγ, as described for thiazolidinediones [39].

Selective PPAR modulators  
Agonists of the nuclear receptor peroxisome proliferator-activated receptor-gamma (PPARγ), such as the current thiazolidinediones (pioglitazone and rosiglitazone), increase insulin sensitivity. Stimulation of PPARγ alters transcription of various genes that mediate metabolic effects of insulin, increase adipogenesis, improve biomarkers of vascular function and exert anti-inflammatory actions (Figure 47.7) [40]. Different PPARγ agonists appear to alter the conformation of the PPARγ binding epitope or adjacent regions of the receptor in slightly different ways such that different co-activators are recruited, varying the selection of genes transcribed and therefore the profile of biologic effects. In addition to further thiazolidinediones, some nonthiazolidinedione PPARγ agonists have been described, including agents that partially modulate PPARγ such as halofenate/metaglidasen and FK614. The differing sets of genes activated by these agonists illustrate the possibility of selective PPARγ modulation to retain desired therapeutic effects and reduce side effects [41]. Selective effects can also be achieved by stimulating genes for particular co-activators such as PGC-1α.

A further approach to selective PPAR modulation has been the development of compounds that bind both PPARγ and PPARα (so-called dual PPARα/γ agonists, or "glitazars"), although previous glitazars (e.g. muraglitazar, tesaglitazar, and aleglitazar) have been discontinued due to side effects. Following evidence that stimulation of PPARα can improve insulin sensitivity, increase fatty acid oxidation, raise thermogenesis, and prevent weight gain, agents have been sought with variable affinities for PPARα, γ, and δ (termed panPPAR agonists or SPPARMs—selective PPAR modulators) [41]. Some modulators of the retinoid X receptor, which forms a heterodimer with each PPAR, have been reported to reduce hyperglycemia, reduce
appetite, increase production of mitochondrial uncoupling proteins, and decrease weight gain; combining these agents with PPARγ agonists is also receiving consideration.

**Vitamins and minerals**

Antioxidants are often depleted in diabetic states, and supplements of antioxidant vitamins have shown small benefits to glycemic control in some studies. Adequate vitamin D concentrations and functioning vitamin D receptors are required for normal activity of islet β cell and of insulin, and diabetes patients are often deficient in circulating vitamin D3 (cholecalciferol). Studies of vitamin D supplementation have so far been equivocal, but it should be borne in mind that correcting a deficiency state is more likely to offer benefit than adding excess in individuals who already have normal values [6,42].

Deficiencies of several minerals, particularly magnesium, chromium, and zinc are often observed in older diabetes patients, and replacement supplements can improve glycemic control in such cases [6]. Magnesium is required for phosphate transfer and kinase reactions, and supplements can enhance insulin action in hypomagnesemic patients. Trivalent chromium supplements can improve glycemic control without increasing insulin concentrations in chromium-deficient patients, and zinc, which is essential for insulin hexamer formation, assists in insulin secretion and action. Lithium, selenium, molybdenum, tungsten, mercury, and cadmium have all been reported to improve glycemic control, although the therapeutic index is generally narrow and toxicity risks have limited clinical investigations. Improved insulin sensitivity with lithium may be offset by decreased insulin secretion resulting in variable effects on glycemic control in T2DM.

Vanadium supplements have improved glycemic control and reduced insulin requirements in type 1 and type 2 diabetes patients, probably due in part to improved insulin sensitivity through inhibition of protein tyrosine phosphatases and suppression of glucose 6-phosphatase. Reduced appetite and a reduced rate of intestinal glucose absorption may also contribute. To limit potential toxicity low-dose peroxovanadium compounds, pervanadates, and organic vanadium complexes have been considered, and it has been noted that the glucose-lowering effects of vanadium salts can persist for weeks after treatment is stopped [6].
Inhibitors of counterregulatory hormones

The major counterregulatory hormones are glucagon and epinephrine which act acutely, and glucocorticoids and growth hormone which have a slower onset and generally more protracted action. They raise blood glucose concentrations by increasing hepatic glycogenolysis and gluconeogenesis, and by impairment of various metabolic effects of insulin. The circulating concentrations of counterregulatory hormones are often inappropriately raised in diabetic states, rendering them a treatment target. However, the counterregulatory response to low glucose concentrations is already delayed or deficient in advanced stages of T2DM, so it is important that interventions should not be so intensive or irreversible that they impede the response at low glucose concentrations.

Glucagon antagonists

Glucagon receptor antagonists have received considerable attention of late [43,44]. Improvements in glycemic control during suppression of glucagon action with glucagon antibodies or peptide antagonists of the glucagon receptor are well known but not particularly suited to routine clinical application. Now there are several small molecule glucagon receptor antagonists that have shown sufficient efficacy in preclinical studies to proceed into clinical trials. Although these are competitive inhibitors of glucagon binding there is also evidence that agents could be developed to uncouple the glucagon receptor from activation of adenylate cyclase [6]. Note that glucagon receptor antagonist may alter other physiologic effects of glucagon such as lipid metabolism, and it is possible that blocking glucagon action could activate some form of feedback to increase glucagon secretion.

As noted earlier, GLP-1 receptor agonists reduce prandial glucagon secretion. The somatostatin analogue octreotide can also suppress glucagon secretion and delay intestinal glucose absorption as well as preventing growth hormone secretion. However, octreotide also inhibits insulin secretion, precluding use in T2DM, although it is potentially useful for glycemic control with insulin in type 1 patients [6].

Glucocorticoid antagonists

Since excess glucocorticoids promote truncal obesity, insulin resistance, and hyperglycemia, tissue-specific inhibitors of glucocorticoid action have been sought to prevent metabolic effects in liver and adipose tissue without significantly interrupting other effects of these steroids. Targeting glucocorticoid receptor binding in liver, for example by conjugating glucocorticoid receptor binding inhibitors to bile salts, has reduced hyperglycemia in animal models [6].

11beta-hydroxysteroid dehydrogenase-1 (11β-HSD1) inhibitors

Cortisol is converted to inactive cortisone in the kidney by 11β-HSD2 and converted back to active cortisol by the reductase activity of 11β-HSD1 which is predominantly expressed in liver and adipose tissue (Figure 47.8). Selective inhibitors of 11β-HSD1, which reduce intracellular cortisol in liver and adipose tissue, have been shown to improve insulin sensitivity, improve glycemic control, reduce body weight, and improve the lipid profile in overweight and obese type 2 diabetes patients [45]. However, avoiding some reduction in circulating cortisol and so preventing a compensatory increase in ACTH has proved challenging. Also, possible effects of inhibiting 11β-HSD1 in islet β cells remain to be established.

Dehydroepiandrosterone (DHEA)

Administration of the adrenal androgen dehydroepiandrosterone (DHEA) and an etiocholanolone metabolite improves glycemic control, reduces insulin resistance and decreases adiposity in obese diabetic animals but this has not been duplicated in clinical studies [46].

Direct modifiers of glucose metabolism

Many substances are known to exert some insulin-like effects that can lower blood glucose concentrations in diabetic animal models. However, most of these substances have other effects that preclude their consideration for long-term clinical use, for example deoxyfrenolicin, vitamin K5, spermine, diamides, and peroxides [6]. Other substances such as okadaic acid and phorbol esters initially imitate certain effects of insulin, but then prevent tissues from responding to insulin. Dichloroacetate and its esters increase glucose oxidation by stimulating pyruvate dehydrogenase, and suppress hepatic glucose production by inhibiting pyruvate carboxylase, but produce glyoxylate and oxolate which can adversely affect neural function.
The antioxidant alpha-lipoic acid, used in some countries to treat diabetic neuropathy, increases insulin sensitivity and improves glucose metabolism, probably due in part to its action as a cofactor for dehydrogenases involved in glycolysis and the Krebs cycle; this has been considered as a possible adjunct to assist glycemic control but is required in high dosages.

**AMP kinase activators**

When the energy levels of cells become depleted, concentrations of adenosine triphosphate (ATP) and creatine phosphate (CP) decline, whereas concentrations of adenosine monophosphate (AMP) become raised. Rising AMP concentrations activate the enzyme adenosine monophosphate-activated protein kinase (AMPK) which promotes energy production by increasing the uptake and oxidation of glucose and fatty acids to replenish ATP (Figure 47.9). AMPK also reduces expression of phosphoenolpyruvate carboxykinase (PEPCK)—a key enzyme for gluconeogenesis, and may suppress cell proliferation. Metformin, thiazolidinediones and adiponectin activate AMPK amongst other actions that contribute to their glucose-lowering effects [47]. Since analogues of AMP such as 5-aminoimidazole-4-carboxamide-1-B-D-ribofuranoside (AICAR) activate AMPK and improve glycemic control in insulin-resistant diabetic animals, further activators of AMPK are under investigation as potential diabetes therapies.

**Glycogen synthase kinase inhibitors**

Glycogen synthesis is impaired in T2DM, due in part to reduced activity of glycogen synthase. Glycogen synthase is normally deactivated by the enzyme glycogen synthase kinase-3 (GSK3), whereas glycogen synthesis proceeds when GSK3 is inactivated by insulin via the PI3K-Akt pathway. Since GSK3 activity is increased in insulin-resistant diabetic states, small molecule inhibitors of GSK3 have been developed which have increased insulin-stimulated glycogen synthesis and lowered blood glucose in insulin-resistant diabetic animals. However, GSK3 has diverse signaling roles affecting neuronal metabolism, protein degradation and apoptosis which present difficulties for specifically targeting glycogen metabolism [6].

**Inhibitors of hepatic glucose production**

Various interventions have been investigated to suppress hepatic gluconeogenesis and/or glycolgenolysis. These need to be reversed rapidly if blood glucose concentrations fall below the normal range so that counterregulatory responses can operate to avoid hypoglycemia [48]. Inhibitors of glycogen phosphorylase have shown glucose-lowering efficacy in diabetic animal models, but there is limited clinical evidence as yet. Glucose-6-phosphatase inhibitors reduce hepatic glucose output and lower blood glucose by preventing the last step in glucose output from both gluconeogenesis and glucoegenesis. However, there is a high risk of hypoglycemia, and cellular accumulation of glucose-6-phosphate increases glycogen deposition and lipogenesis which predispose to fatty liver. The step in gluconeogenesis before glucose-6-phosphate is the formation of fructose-6-phosphate from fructose-1,6-bisphosphate, mediated by fructose-1,6-bisphosphatase (F16Pase). Inhibitors of F16Pase reduce hepatic glucose output and may cause a compensatory increase in glycolgenolysis which could help to guard against hypoglycemia.

**Modifiers of lipid metabolism**

Dyslipidemia in T2DM typically involves raised concentrations of very low-density lipoprotein (VLDL) triglyceride and nonesterified fatty acids (NEFAs) which contribute to insulin resistance and hyperglycemia [21]. Raised concentrations of fatty acids and their metabolites impair cellular insulin signaling pathways and alter fuel selection through the glucose–fatty acid (Randle) cycle. Thiazolidinedione PPARγ agonists act in part by altering fatty acid metabolism, and lipid-lowering fibrates, which are PPARδ agonists, may improve glycemic control in some patients. Other agents that lower plasma triglycerides or NEFA concentrations have been shown to acutely improve glycemic control in T2DM, but effects have not been consistent and often not sustained [6].

Inhibiting fatty acid oxidation interrupts the energy supply for hepatic gluconeogenesis and enhances the use of glucose as a source of energy in skeletal muscle. Most inhibitors of fatty acid oxidation act by inhibiting carnitine palmitoyltransferase-1 (CPT-1), the rate-limiting enzyme for transfer of long-chain fatty acyl-CoA into the mitochondria. These have effectively lowered glucose concentrations but carry a high risk of hypoglycemia [6].
**Sodium-glucose co-transporter inhibitors**

The sodium-glucose co-transporter-2 (SGLT2), located almost exclusively in the proximal convoluted tubules, is a high capacity co-transporter responsible for reabsorption of most of the glucose filtered by the renal glomeruli. Partial competitive reversible inhibition of this co-transporter reduces glucose reabsorption, enabling excess glucose to be eliminated in the urine (Figure 47.10). The mechanism is independent of insulin, allowing inhibitors of SGLT2 to be used alone or in combination with other therapies to reduce hyperglycemia at different stages in the natural history of T2DM, provided there is adequate renal function [49]. The extra glucosuria typically accounts for 50–90 g glucose cleared daily (equating 200–360 kcal) depending on the extent of the hyperglycemia and renal function. Thus, in addition to reducing hyperglycemia, SGLT2 inhibitors can assist body weight reduction. Since SGLT2 inhibition does not stimulate insulin secretion or interrupt the counterregulatory system, there is low risk of hypoglycemia unless used with agents that carry their own inherent risk of hypoglycemia. While glucosuria increases the risk of urinary and genital infections, the osmotic diuresis generated by the glucosuria may benefit blood pressure control as seen with a mild diuretic. Four members of the SGLT2 inhibitor class (dapagliflozin, canagliflozin, empagliflozin, and ipragliflozin) have recently been introduced.

SGLT1 is a low-capacity co-transporter also located within the proximal tubules but distally to SGLT2. SGLT1 can reabsorb small amounts of glucose that evade reabsorption by SGLT2, and blockade of SGLT1 can support the glucosuric effect. SGLT1 is also the co-transporter for glucose absorption by the small intestine, and its inhibition can delay and slow glucose absorption. Hence agents with inhibitory effects on SGLT2 and SGLT1 are in development with pharmacokinetic properties selected to avoid malabsorption and passage of glucose into the large bowel. It is also noted that renal reabsorption of low glucose concentrations by SGLT1 may help to provide a buffer against the risk of hypoglycemia.

**Sirtuins**

Sirtuin enzymes are nicotinamide-adenine-dinucleotide (NAD)-dependent deacetylases and/or ADP-ribosyltransferases. Their effects include altering gene transcription through chromatin silencing, and they mediate metabolic responses that...
resemble chronic caloric restriction. Sirtuin sirt1 is widely expressed in mammalian tissues including liver, muscle and fat, and promotes mitochondrial biogenesis, increasing thermogenesis and reducing susceptibility to weight gain, diabetes, and cardiovascular disease. AMPK enhances sirt1 activity, resulting in increased activity of downstream transcription factors such as PGC-1α and the forkhead protein FOXO1 which increase energy production. Sirt1 in pancreatic β cells may also facilitate insulin secretion. Several small molecule activators of sirt1 such as the phytophenol resveratrol have been shown to increase mitochondrial density, and improve insulin sensitivity, glycemic control, and eNOS-dependent vasorelaxation in animal models, providing a template for other activators of sirt1 [50].

Conclusions

The multivariable etiology, pathogenesis, progression and morbidities of T2DM require the development of a variety of different treatments that can be used to address the different adverse features of the disease as it advances. There are new types of agents at various stages of development to improve the function of the pancreatic islets, enhance insulin action, reduce glucose production, facilitate glucose metabolism, or reduce the impact of internal factors that act counter to the reinstatement of normal glucose homeostasis. Despite the many initiatives under investigation there is still an evident paucity of prospective interventions to prevent or reverse the loss of functional β-cell capacity and responsiveness, and a need for additional measures to address the diverse consequences of insulin resistance.

References

New drugs for the treatment of diabetes mellitus


29 Popov D: Novel protein tyrosine phosphatase 1B inhibitors: interaction requirements for improved intracellular efficacy in type 2 diabetes mellitus and obesity control. *Biochemical and Biophysical Research Communications* 2011; 410:377–381.


CHAPTER 48

Incretin-based therapies

Michael A. Nauck1 and Juris J. Meier2
1 Diabeteszentrum Bad Lauterberg, Bad Lauterberg im Harz, Germany
2 Division of Diabetes and Gastrointestinal Endocrinology, University Hospital St. Josef-Hospital, Ruhr-University Bochum, Gudrunstr, Bochum, Germany

Key points

• Incretin-based medications make use of the glucose-lowering activity of the gut incretin hormone glucagon-like peptide-1 (GLP-1), which stimulates glucose-dependent insulin secretion, suppresses glucagon secretion, decelerates gastric emptying, reduces appetite and food intake, and may have other actions in the nervous and cardiovascular systems.

• Incretin mimetics (GLP-1 receptor agonists) are peptides that need to be injected subcutaneously (such as insulin) and have a much longer half-life than native GLP-1, so can be injected at intervals between twice daily (exenatide, unretarded), once daily (liraglutide), or once weekly (exenatide once weekly, and novel compounds currently under development).

• DPP-4 inhibitors preserve GLP-1 secreted from endogenous sources in its intact, biologically active state, so that for the same amount released from the intestines more GLP-1 receptor stimulation takes place, thereby increasing insulin secretion relative to ambient levels of glycemia, and suppressing glucagon secretion.

• Incretin-based medications have the advantage of not causing hypoglycemic episodes due to their glucose-dependent mechanism of action on insulin and glucagon secretion. Nevertheless, hypoglycemia can be caused by concomitant medications (insulin, sulfonylureas) and this risk is increased by incretin-based antidiabetic medications because of their glucose-lowering activity.

• Incretin mimetics induce some weight loss in the average patient, but there is marked heterogeneity between patients regarding their susceptibility towards weight loss induced by GLP-1 receptor stimulation. Nevertheless, this is attractive for patients who have unsuccessfully tried to lose weight on medications that tend to increase body weight, such as sulfonylureas, meglitinides, thiazolidinediones, or insulin.

Biologic actions of incretin hormones in patients with type 2 diabetes mellitus

Incretin hormones are gut peptides released after nutrient ingestion which have the ability to augment insulin secretory responses, provided there is a permissive rise in glycemia, as is typical after meals. The two well-defined incretin hormones are glucose-dependent insulinotropic polypeptide (formerly called gastric inhibitory polypeptide; GIP) and glucagon-like peptide-1 (GLP-1). Their physiologic role is best exemplified by looking at the difference in insulin secretory responses after an oral glucose challenge in healthy subjects compared to an “isoglycemic” intravenous (i.v.) glucose infusion, which provides an identical rise in glycemia (i.e., an identical glycemic stimulus to the endocrine pancreas). Nevertheless, much more insulin (and C-peptide) is released in response to the oral stimulus, which leads to the release (from the gut) and insulinotropic action (on the endocrine pancreas) of the incretin hormones. This phenomenon is called the incretin effect and is responsible for two-thirds of insulin secretory response following a standard (75 g) oral glucose challenge. For more details on the physiology and pathophysiology of the incretin system, see Chapter 11 on incretin physiology in health and disease. GIP was discovered earlier, probably contributes most of the incretin effect in healthy human subjects, but is inactive as an insulinotropic agent in patients with T2DM, because type 2 diabetic β cells show very weak responses to GIP only. This explains why patients with T2DM have a reduced or even absent incretin effect. GLP-1 probably makes a minor physiologic contribution to the incretin effect (among other reasons, because gastric emptying is delayed, preventing a permissive glycemic rise), but retains much of its insulinotropic activity in T2DM. Thus, it has been possible to elicit an insulin secretory response to exogenous GLP-1 at low pharmacologic concentrations in patients with T2DM, and it has also been possible to normalize fasting glucose in hyperglycemic patients with T2DM. Such proof-of-principle experiments demonstrate the therapeutic potential of GLP-1 receptor stimulation to lower glucose in these patients. On the other hand, GLP-1 is rapidly degraded and inactivated by dipeptidyl peptidase-4 (DPP-4), a ubiquitous protease, and in addition, is rapidly eliminated via the kidneys.
such that the half-life of intact, biologically active GLP-1 (7-36 amide) is only approximately 1.5 min [1,2].

**GLP-1 as a parent compound for incretin-based antidiabetic drugs**

1. GLP-1 can augment glucose-induced insulin secretion in healthy subjects and in patients with T2DM.
2. GLP-1 suppresses glucagon secretion at glucose concentrations ranging from the normal fasting range to hyperglycemia. At hypoglycemic blood glucose concentrations, when glucagon is needed as an important counterregulatory hormone, GLP-1 no longer suppresses glucagon secretion—the counterregulatory response remains normal even in the face of high, pharmacologic concentrations of GLP-1.
3. GLP-1 delays gastric emptying. This leads to a slower entry of nutrients into the circulation, and smaller rises in glucose and triglyceride concentrations after meals.
4. GLP-1 in the general circulation has access to brain areas involved in the regulation of satiety, appetite, food intake, and energy expenditure. GLP-1 increases satiety and the feeling of fullness, reduces appetite and the desire to eat, and decreases food (energy) intake. These are the prerequisites for weight loss in the longer term. These actions match the typical therapeutic challenges in patients with T2DM, namely,
   - defective insulin secretion due to a reduced mass and function of β cells;
   - hyperglucagonemia leading to exaggerated hepatic glucose output in the fasting state and after meals;
   - exaggerated hyperglycemia after eating; and
   - obesity [2].

In addition to these core biologic activities, the following biologic activities have been described:

1. Improvements in (pro-)insulin biosynthesis, induction of neo-genesis and/or proliferation of β cells, and prevention of apoptosis induced by a variety of toxic factors (hydrogen peroxide, cytokines, and so on). Most of these experiments have been performed in young rodents and cell lines. It has not been as easy to demonstrate similar findings in older animals. It is an open question whether such β-cell-preserving effects apply to patients with T2DM treated with incretin-based medications;
2. Neuroprotective actions in the (central) nervous system;
3. Cardiovascular actions at the level of the endothelium, leading to vasodilatation, and at the level of the ischemic or failing heart, leading to reduced myocardial necrosis in the case of induced myocardial ischemia, and increased oxygen and glucose uptake and utilization and improved left ventricular function in the case of congestive heart failure [2].

**Development of GLP-1 receptor agonists (incretin mimetics)**

For therapeutic purposes, compounds with the ability to activate GLP-1 receptors, but with a much longer half-life than the parent compound GLP-1 were needed. Ideally, such compounds should be resistant to proteolytic degradation and inactivation by DPP-4. Two approaches were taken: (i) natural compounds with GLP-1 receptor agonistic properties were sought, and (ii) GLP-1 was intentionally modified with the aim of preserving a high degree of homology with the parent compound, but prolonging the half-life to allow for reasonable inter-injection intervals [2].

**Exenatide**

Exenatide (Byetta®) is synthetic exendin-4, a peptide identified in the saliva from Heloderma suspectum (the “Gila monster”), a lizard from Arizona. Similarities in the amino acid sequence to GLP-1 prompted examinations demonstrating exenatide to be a potent GLP-1 receptor agonist. Exenatide is a longer peptide molecule than GLP-1, and displays approximately 53% sequence homology. Pharmacokinetic examinations in human subjects indicate a half-life of approximately 2–3 hours. After a single subcutaneous injection, plasma levels peak after 2 h and remain elevated for 6–7 h [3]. A phase 2 study indicated that three injections before the main meals were no more efficient at reducing HbA1c than two injections before breakfast and dinner. Almost all studies since have employed two injections per day. The injection should be timed within 60 min before ingesting the respective meal. Fifteen minutes prior to eating is a reasonable recommendation. Exenatide is administered using a pen device. Pens are available that deliver 5 or 10 μg per dose. The recommended starting dose is 5 μg twice daily. After 4 weeks (when exenatide from the first pen has been used up), patients can change to another pen that delivers 10 μg per dose. The maintenance dose for most patients is 10 μg twice daily, unless adverse events suggest continuing at the lower dose because it is better tolerated, or in cases of moderate renal functional impairment (eGFR 30–50 mL min⁻¹), making a dose reduction necessary.

**Liraglutide**

Liraglutide (Victoza®) was the first “designer molecule” developed as a GLP-1 receptor agonist. The core peptide has only been slightly modified (in comparison to GLP-1 (7-36 amide), the homology being 97%). A lysine residue in amino acid position 34 had to be changed to an arginine residue, because a side chain with a free fatty acid (FFA) moiety was to be attached to another lysine residue in a selective manner. This FFA side chain (similar to insulin detemir) mediates binding to albumin, the noncovalent liraglutide-albumin conjugate forming a reservoir, from which liraglutide can be liberated to diffuse into tissues and bind to receptors. Although the amino acids at the amino-terminus have not been changed relative to the native GLP-1 peptide, DPP-4 hardly attacks liraglutide, most likely because the FFA side chain can cause steric hindrance. The half-life of liraglutide is approximately 13 hours. With one subcutaneous injection, plasma concentrations remain elevated after 24 h, and with repeated, once-daily subcutaneous injections, a steady state of plasma concentrations is reached after approximately 5 days [4]. Within a 24-h period,
plasma concentrations then fluctuate by approximately 25%. Thus, liraglutide concentrations are permanently elevated with once-daily injections, including the overnight period, where stimulation of insulin secretion and suppression of glucagon secretion lead to a substantial reduction in fasting plasma glucose. Such a prominent effect on fasting plasma glucose is not seen with shorter-acting GLP-1 receptor agonists such as exenatide and lixisenatide. Liraglutide injections do not need to be timed with meals. Patients should choose a time point during the day when they can most conveniently deliver the liraglutide injection every day (±1 h). Liraglutide therapy is started at a subcutaneous dose of 0.6 mg d$^{-1}$. After one week the dose can be raised to 1.2 mg d$^{-1}$. There is an option to further increase the dose to 1.8 mg d$^{-1}$ but stepping up the dose from 1.2 to 1.8 mg d$^{-1}$ is not expected to greatly improve glycemic control [5]. However, slightly more weight loss may be expected at the higher dose. In fact, weight loss trials in nondiabetic obese subjects have utilized up to 3.0 mg liraglutide per day, improving weight loss over lower doses.

**Exenatide once weekly**

Exenatide is now available in a new preparation (Bydureon®) developed for slow and prolonged release from microspheres that encapsulate the compound in a polymer that slowly dissolves in the subcutaneous tissue and releases the peptide. Exenatide once weekly is available only at one dose, 2 mg. For subcutaneous injections, exenatide powder needs to be re-suspended in a single-use device. Since this suspension is more viscous than an aqueous solution, an 8 mm 23-gauge needle is used as part of this package. With once-weekly subcutaneous injections, it takes 6–10 weeks before a steady state of concentrations is reached [6], which thereafter is maintained with once-weekly injections. After discontinuing treatment, it takes several weeks before all previously injected exenatide is absorbed and eliminated.

**Agents under development**

Albiglutide is composed of two molecules of modified GLP-1 attached to one molecule of human albumin. Despite the much larger molecular size (compared to GLP-1, exenatide, and liraglutide), albiglutide is able to stimulate the GLP-1 receptor. However, because of the greater molecular size, the molecule may not fully access the central nervous system. Albiglutide has a half-life of approximately 5 days in humans and needs to be injected once weekly. Attempts to prolong the interval between injections to 2 or even 4 weeks led to fluctuations in glycemic control throughout the inter-injection interval, probably prompted by fluctuations in plasma concentrations of albiglutide. Albiglutide has not yet been approved. Most likely, once-weekly doses of 30 mg (for an initial treatment of 6 weeks) and 50 mg (for maintenance treatment) will be used [7].

Lixisenatide is a derivative of exenatide, which has so far not been approved for the treatment of T2DM. It has a similar short action to exenatide, but lixisenatide has been developed for once-daily subcutaneous dosing [7]. Thus, it has limited ability to lower fasting glucose in hyperglycemic patients, but like exenatide, it can profoundly affect postprandial glycemic excursions through its ability to retard gastric emptying even with long-term use. With these properties, it is likely to be used mainly as an adjunct to insulin therapy.

Dulaglutide is a fusion molecule of GLP-1 and an immunoglobulin Fc fragment. It is being developed for administration once weekly [8]. Protracted effects of PEGylated GLP-1 have been described in preclinical studies.

Small molecules activating the GLP-1 receptor have been found that may prove that oral agents can be developed in addition to the GLP-1-agonistic peptides that are currently used as incretin mimetics.

**Efficacy of incretin mimetics as glucose-lowering agents (reduction in HbA1c)**

For diabetes medications, the key feature of incretin mimetics is their ability to lower/control glucose, a prime factor in preventing long-term complications induced by long-standing hyperglycemia.

**Monotherapy**

Exenatide at 10 μg twice a day is able to reduce HbA1c by 0.9 % (0.2% with placebo) in patients with T2DM not concomitantly treated with other (oral) glucose-lowering agents [9]. Liraglutide has been proven to lower HbA1c by 1.1% in monotherapy [10]. Exenatide once weekly has been studied in populations including patients with T2DM not previously or concomitantly treated with other glucose-lowering agents. The results for these subpopulations have not been reported separately. The main use of GLP-1 receptor agonists is as add-on to oral glucose-lowering therapy, since guidelines uniformly recommend metformin as the first choice for initial drug therapy. However, exenatide (unretarded), liraglutide, and exenatide once weekly can be used in drug-naïve patients in monotherapy, if metformin and other agents are contraindicated or cannot be tolerated.

**Add-on to oral glucose-lowering agents**

For injectable agents like GLP-1 receptor agonists, the logical place in the treatment algorithm is as add-on to oral antidiabetic agents as an alternative to insulin treatment, or to other oral glucose-lowering drugs in cases where these agents cause adverse effects. Exenatide (unretarded), liraglutide, and exenatide once weekly have been studied against a background of metformin, sulfonylureas, a combination of metformin and sulfonylureas, thiazolidinediones, and a combination of metformin and thiazolidinediones. In many studies, mixed populations encompassing, for example, drug-naïve patients, metformin-treated patients, and patients treated with a combination of metformin and sulfonylureas, have
Incretin-based therapies

Baseline HbA1c [%]:

Baseline BMI [kg/m²]:

Trial duration [weeks]:

30 26 15 16 13 12

Exenatide twice daily

Dose:

40 30 20 10 0

Exenatide once weekly

Albiglutide

Lixisenatide

Dulaglutide

Placebo

Liraglutide

8.3 8.0 7.5 7.3

Δ HbA1c [%]

Δ Body weight [kg]

Δ Fasting plasma glucose [mmol/L]

Proportion with hypoglycemia [%]

Nauck et al. 2009

DeFronzo et al. 2005

Kim et al. 2007

Rosenstock et al. 2009

Ratner et al. 2010

Grunberger et al. 2012

Figure 48.1 Essential results from clinical trials with different GLP-1 receptor agonists compared to placebo treatment, both administered on a metformin background. Effect on HbA1c (a), fasting plasma glucose (b), body weight (all Δ versus baseline) and proportion of patients reporting at least one episode of hypoglycemia (d). Trials can be identified by first author and year of publication. Baseline HbA1c (a), trial duration (b), baseline BMI (body mass index) (c), and dose as well as the typical interval between two injections (d) are also indicated. For references, see [5,7,8,23,49,50].

been examined without reporting the treatment results in subgroups. Generally speaking, the ability of GLP-1 receptor agonists to control glycemia does not depend to a great extent on background treatment with oral glucose-lowering agents (Figure 48.1).

Head-to-head studies comparing incretin mimetics have shown important differences: long-acting agents (i.e., those that lead to a constantly elevated plasma drug concentration throughout the 24-h period, such as liraglutide and exenatide once weekly), show lower overnight and fasting glucose concentrations, resulting in a greater reduction in HbA1c; whereas with short-acting agents, peak-to-trough levels are observed between injections (unretarded exenatide, lixisenatide). Short-acting agents exert only limited effects on fasting glucose, because with the typical injection regimen, overnight drug levels are low. Thus, unretarded exenatide has been shown to be less effective in controlling HbA1c than liraglutide [11] or exenatide once weekly [6].

Liraglutide at 1.8 mg d⁻¹ has also been compared to exenatide once weekly (2 mg per week) and albiglutide (30 mg once weekly followed by 50 mg once weekly after week 6). In both comparisons, liraglutide treatment was associated with a superior reduction in HbA1c [12,13], although the difference was a mere 0.2–0.3 % and thus of negligible or questionable clinical relevance. Nevertheless, liraglutide at a once-daily injection regimen appears to be the most potent GLP-1 receptor agonist currently available for the treatment of T2DM. Explanations for this may be:

1. that the trials were too short for the longer acting incretin mimetic to catch up with the more pronounced initial effectiveness of liraglutide, which reaches a steady state after approximately 5 days rather than several weeks;
2. only single doses of each agent were compared, and the dose–response relationships have not been characterized to the degree necessary to unequivocally define optimum doses for each GLP-1 receptor agonist. Thus, conclusions cannot be drawn that any agent is more effective than another compound, and should be limited to the doses employed in the clinical trials;
3. the possibility cannot be excluded that larger inter-injection intervals lead to fluctuations in drug levels and associated effects between two injections, with weaker effects at times characterized by relatively lower drug concentrations.

Comparison to insulin treatment

For all available GLP-1 receptor agonists, clinical trials have been carried out to compare their glycemic control effects with those of insulin regimens. Exenatide (unretarded) has been compared to both premixed insulin [14,15] and the long-acting insulin analogue insulin glargine; liraglutide [16] and exenatide once
weekly [17] have been compared to insulin glargine. In all comparisons, the incretin mimetic was at least as effective in lowering HbA1c as the insulin regimen. In some trials, the insulin preparation was more effective at reducing fasting glucose, suggesting a greater postprandial mode of action by the incretin mimetic, but this was not the case when comparing liraglutide with insulin glargine [16]. Since all GLP-1 receptor agonists are used in a standard dose, while insulin always needs to be titrated, the question arose whether weaknesses (inertia) in the titration process might explain the relatively good performance of the incretin mimetics. Some of the studies achieved average insulin doses similar to those in treat-to-target trials, contradicting this assumption. Most likely, in clinical practice, similar limitations apply to the titration process of insulin. Thus, GLP-1 receptor agonists are similarly potent as insulin in helping patients with T2DM to achieve their HbA1c targets (Figure 48.2).

**Combination with basal insulin**
Short-acting GLP-1 receptor agonists considerably decelerate gastric emptying even with long-term use [18]. Therefore, it is probably not wise to combine them with rapid-acting insulin injected before meals, because this combination is likely to provoke hypoglycemia as the insulin will peak before the nutrients enter the circulation. However, one prominent “weakness” of short-acting incretin mimetics is their relatively minor effect on fasting glucose, while their obvious strength is the control of postprandial glycemic rises, mostly mediated by the retardation in gastric emptying which persists with continued use of exenatide and lixisenatide. On the other hand, regimens that rely on long-acting insulin (mostly in combination with oral antidiabetic agents, preferably metformin) can control fasting glycemia very well, but are often associated with relatively large postprandial glycemic increments. Thus, a combination of a long-acting insulin such as insulin glargine or insulin detemir with a short-acting incretin mimetic like exenatide or lixisenatide can provide good glycemic control in the fasting and postprandial state. Comparative trials to intensified insulin regimens (multiple daily injections) are not available, but insulin glargine in combination with exenatide (unretarded) leads to weight loss, and the addition of exenatide to insulin glargine did not increase the number or severity of hypoglycemic episodes [19]. Benefits of combining liraglutide with insulin detemir have also been reported [20]. This combination may be among the most efficient regimens to treat long-standing T2DM in patients in whom other regimens are more likely to fail.

---

**Figure 48.2** Essential results from clinical trials with different GLP-1 receptor agonists compared to insulin treatment, both administered on a background of oral antidiabetic agents. Effect on HbA1c (a), fasting plasma glucose (b), body weight (all Δ versus baseline) and proportion of patients reporting at least one episode of hypoglycemia (d). Trials can be identified by first author and year of publication. Baseline HbA1c (a), trial duration (b), baseline BMI (body mass index) (c), and dose as well as the typical interval between two injections (d) are also indicated. N.r., not reported; Δ, the bar indicates the difference between insulin and exenatide treatment. For references, see [14–17,51,52].
Inability to provoke hypoglycemia

GLP-1 even in high doses and GLP-1 receptor agonists at low glucose concentrations cannot elicit insulin secretory responses or other glucoregulatory activities potent enough to cause hypoglycemia. The reason is the inability of GLP-1 receptor stimulation to close ATP-dependent potassium channels in β cells, a step necessary to initiate insulin release. In line with this physiologic explanation, hypoglycemia rates in clinical trials involving incretin mimetics have always been low, at similar levels to placebo or metformin. Severe episodes of hypoglycemia have been observed extremely rarely under such conditions, most trials reporting no such episodes. This changes, however, as soon as incretin mimetics are combined with other drugs which have the potential to induce hypoglycemia, namely sulfonylureas, meglitinides (which have the potential to close ATP-dependent potassium channels at any ambient level of glycemia), or insulin. Under such circumstances, the rate of hypoglycemia usually increases with treatment using incretin mimetics, because of their additional glucose-lowering effect. Since providing glycemic control with a low or absent risk for hypoglycemic episodes is an important goal, potential benefits need to be weighed against the potential to provoke hypoglycemia before combining incretin mimetics with sulfonylureas or meglitinides. The addition of short-acting incretin mimetics to treatment with long-acting insulin does not seem to increase the risk of hypoglycemia substantially [2].

Deceleration of gastric emptying

GLP-1 and GLP-1 receptor agonists decelerate gastric emptying through a mechanism involving the parasympathetic nervous system (afferent and efferent vagus branches). Acutely, GLP-1 at doses that do not cause gastrointestinal side effects, can lead to a standstill of gastric emptying. Lower doses just retard gastric emptying, by interrupting the transpyloric delivery of nutrients, thus delaying the absorption of nutrients and their entry into the circulation. This is the main reason that incretin mimetics profoundly limit postprandial glycemic rises in glucose, and also in triglyceride concentrations. It is likely that all incretin mimetics delay gastric emptying on first exposure [21]. With prolonged treatment (i.e., after days to weeks), there appears to be tachyphylaxis regarding the ability to delay gastric emptying with long-acting GLP-1 receptor agonists, because the typical dosing regimen leads to permanently elevated plasma concentrations of the compound with no nadirs close to zero concentrations. This description is true for liraglutide, exenatide once weekly, and albiglutide. Only short-acting incretin mimetics like exenatide (unretarded) and lixisenatide preserve their ability to retard gastric emptying with prolonged treatment, because the exposure to the agent is intermittent, with low drug levels close to zero concentrations in between injections, which are apparently necessary to maintain the full responsiveness to this action following GLP-1 receptor stimulation [22]. It is important to mention that other activities induced by GLP-1 receptor stimulation are not subject to tachyphylaxis. This applies to the stimulation of insulin secretion, the suppression of glucagon secretion, and the reduction in appetite and food intake. The only other example of reduced effectiveness with prolonged exposure to GLP-1 receptor agonists is the reduced probability of gastrointestinal adverse events with longer treatment duration.

Inducing weight loss

Incretin mimetics, on average, induce some weight loss. This weight loss is in the order of 2–4 kg for the average patient, and appears to be maintained as long as GLP-1 receptor agonist treatment is continued. At the individual patient level, the weight response is heterogeneous, with some patients losing more than 10 kg, while others do not lose weight at all or may even gain some weight. What determines the individual responses is not known. Generally speaking, weight loss, even if it is moderate for most patients, is an attractive component of the spectrum of clinical effects induced by treatment with incretin mimetics, considering that other antidiabetic medications like sulfonylureas, meglitinides, thiazolidinediones, and insulin typically cause patients to gain body weight.

The signaling pathway that leads to a reduction in appetite and food intake when patients are treated with incretin mimetics involves central nervous system areas in the hypothalamus responsible for the regulation of energy intake and expenditure. Whether GLP-1 receptor agonists cross the blood–brain barrier or interact with receptors in parts of the central nervous system that are devoid of a blood–brain barrier (“subfornical organs”), or whether the primary interaction is with GLP-1 receptors on afferent parasympathetic (vagal) branches is not known [2].

Theoretically, reduced food intake could also be due to the fact that incretin mimetics retard gastric emptying. It is obvious that if the stomach is still partially filled from a previous meal, this could lead to reduced appetite and food intake. However, although theoretically this sounds plausible, no direct evidence has been provided that this mechanism is active and contributes to weight loss during treatment with GLP-1 receptor agonists.

Short-acting (e.g. exenatide, unretarded) and long-acting (e.g. liraglutide, exenatide once weekly) GLP-1 receptor agonists appear to have a similar potential to induce weight loss, based on head-to-head comparisons between these agents. Since treatment with short-acting incretin mimetics is characterized by a persistent deceleration of gastric emptying, even with long-term treatment, whereas long-acting GLP-1 receptor agonists lose this activity through tachyphylaxis, this speaks against a major contribution of delayed gastric emptying to weight loss. Alternatively, other mechanisms may compensate for the varying contribution of delayed gastric emptying.

In head-to-head comparisons of liraglutide with exenatide once weekly [11] and with albiglutide [13], weight loss was
more pronounced with liraglutide than with the once-weekly preparations. This could be the consequence of an overall lesser effect, perhaps related to the doses used in these particular studies. In the case of albiglutide, the difference was substantial and raises the question whether the ability of albiglutide to trigger the mechanisms leading to weight loss is reduced in comparison with other incretin mimetics, possibly related to its larger molecular size, which could be responsible for reduced access to the central nervous system.

The potential to induce weight loss is an important factor driving the prescription of incretin mimetics. Whereas the use of antidiabetic medications that promote weight gain often leads to demotivation amongst patients whose efforts to adopt a healthy lifestyle, including avoidance of excess caloric intake, are unsuccessful, supporting weight reduction pharmacologically hopefully will motivate patients to eat healthy amounts of food for prolonged periods of time. Nevertheless, the effects of patient education supporting healthy eating in connection with the initiation of treatment with incretin mimetics have not received the emphasis that they probably deserve. Most likely, patients could be trained to become sensitive, with the help of GLP-1 receptor agonists, to the signals that should make them stop eating, once satiety and fullness have reached a certain level after meals.

### Gastrointestinal adverse events

GLP-1 and all GLP-1 receptor agonists can cause nausea, vomiting, diarrhea and other gastrointestinal adverse events. Such side effects are thought to be dose-related. Often, they are thought to be mediated by the well-known deceleration of gastric emptying, in the sense that in the absence of propulsive movement of gastric contents, distension and symptoms will be the consequence. More likely, these adverse events are mediated by a direct interaction of GLP-1 and the GLP-1 receptor agonist with target areas of the central nervous system, and thus, can occur in fasting subjects and after food intake.

Uniformly, with all GLP-1 receptor agonists, two phenomena have been observed. (1) The percentage of patients affected by nausea, vomiting, and/or diarrhea will be highest upon initial exposure, gradually declining over days or weeks, until just a small percentage of patients remain, in whom persistent gastrointestinal side effects are observed [5,6,23,24]. Some patients will not tolerate this and will discontinue treatment with GLP-1 receptor agonists. This applies to approximately 3–5% of the type 2 diabetic population. (2) If the dosing schedule is such that the initial exposure to the GLP-1 receptor agonist is lower, and then is gradually increased over time, higher doses can be tolerated in the long term. Therefore, exenatide (unretarded) treatment is usually initiated at 3 μg per dose and later increased to 10 μg per dose. Liraglutide treatment is usually initiated at 0.6 mg d⁻¹ and, after one week, increased to 1.2 mg d⁻¹ (perhaps later followed by a further increase to 1.8 mg d⁻¹). With exenatide once weekly, even when starting the injections at the full dose (2 mg per week), the slow absorption and elimination take care of a slowly rising plasma concentration of the agent.

In head-to-head comparisons of different GLP-1 receptor agonists, the incidence of nausea, vomiting, and/or diarrhea can differ between the compounds. As a rule, long-acting GLP-1 receptor agonists providing permanent exposure to elevated drug levels will produce less gastrointestinal side effects, especially during long-term treatment [6,11]. Certainly, the prevalence of such side effects will be reduced over time, unlike the more persistent levels found with short-acting GLP-1 receptor agonists (characterized by a rapid alternation of periods with high and low drug levels). More specifically, exenatide (unretarded) and liraglutide produce similar levels of gastrointestinal side effects initially, but with liraglutide, these levels decline over time. Exenatide once weekly and albiglutide produced less gastrointestinal side effects than did liraglutide, but they were also less effective in these trials with respect to glycemic control and weight loss, so the doses used for each agent may not have been absolutely equivalent. Differences regarding the access to the relevant brain areas may also play a role.

### Antibody formation

All GLP-1 receptor agonists available for the treatment of T2DM are peptides, which typically are administered by subcutaneous injection. As with other peptides (e.g. insulin), even when they are identical or similar to those that are endogenously produced, repeated injection into the subcutaneous compartment can lead to antibody formation by the immune system. Such antibodies can reach different titres (i.e., amounts of antibodies produced) and different affinities (i.e., avidity, with which the antigen is bound), and may or may not bind the peptide in a manner that neutralizes its biologic activity. Generally speaking, the likelihood of prompting an antibody response should depend on the homology of the GLP-1 receptor agonist with endogenously produced peptides, in this case GLP-1. Exenatide has 53% amino acid sequence homology to GLP-1. With exenatide (unretarded), approximately 45% of patients develop antibodies [11]. This number increases to 60–75% with the extended-release preparation (exenatide once weekly) [6]. A few percent develop high-titre antibodies, and there is suggestive evidence that in those with high-titre antibodies, there is less clinical effectiveness of exenatide treatment [6]. Liraglutide has 97% amino acid sequence homology with GLP-1. Up to 8% of liraglutide-treated patients develop antibodies, so far without a hint that the clinical effectiveness of this treatment might be compromised by antibody formation [25].

### Development of inhibitors of dipeptidyl peptidase-4 (DPP-4)

GLP-1 (like the partner incretin hormone, glucose-dependent insulinotropic polypeptide; GIP) is rapidly degraded by the
ubiquitous endopeptidase dipeptidyl peptidase-4 (DPP-4). The original observation after the development of specific assays that enabled measurement of intact and biologically active GLP-1, was that even with a constant intravenous infusion of GLP-1 in patients with T2DM, only one sixth of the GLP-1 concentration would remain in the intact state, while the majority had already been degraded. This proved (a) the rapidity and (b) the effectiveness of the interaction between GLP-1 and DPP-4, and suggested that interference with DPP-4 activity (inhibition) might preserve endogenously produced GLP-1 in the intact and biologically active form, thus prompting a glucose-lowering effect of DPP-4 inhibitor treatment. Indeed, studies delineating the mode-of-action of DPP-4 inhibitors show the inhibition of DPP-4 enzyme activity (typically, by 85% or more when determined in serum samples, i.e., measuring the soluble form of DPP-4), a rise in intact, biologically active GLP-1 after oral glucose or meal stimulation by a factor of 2–3, a slight light reduction fasting and a more pronounced reduction in postload glycemıc excursions, along with unchanged insulin responses and suppressed glucagon responses. This pattern is exactly what can be expected from higher endogenous GLP-1 levels, that is, similar insulin responses at lower glycemıc excursions prove insulınotropic activity, and reduced glucagon responses demonstrate glucagonostatic activity [26].

It may be viewed as a surprise that with near-complete inhibition of DPP-4, the intact, biologically active GLP-1 concentrations with DPP-4 inhibition only double or at most triple, since typically, only approximately 15% of GLP-1 are in the intact state (thus, theoretically, allowing for a sixfold rise when the process of degradation/inactivation is interfered with almost completely). The explanation is a feedback inhibition of L-cell (GLP-1) secretion elicited by higher concentrations of intact GLP-1 prompted by DPP-4 inhibition [27].

**Sitagliptin**

Sitagliptin is (2R)-4-Oxo-4-[3-(trifluoromethyl)-5,6-dihydro[1,2,4]triazolo[4,3-α]pyrazin-7(8H)-yl]-1-(2,4,5-trifluorophenyl)butan-2-amine and has a triazolopiperazine or β-amino acid-based structure [28–30]. It is a competitive inhibitor of DPP-4 and has a long half-life allowing for once daily dosing (100 mg in patients with normal renal function). Details of its pharmacologic properties are shown in Table 48.1. Sitagliptin is the DPP-4 inhibitor that has been approved for the longest time (since 2006).

**Vildagliptin**

Vildagliptin is 1-[(3-hydroxy-1-adamantyl) amino]acetyl]-2-cyano-(S)-pyrrolidine with a cyano-pyrrolidine structure [29–31]. Despite its relatively short half-life (Table 48.1), its interaction with DPP-4 is of longer duration, because it forms a covalent bond with DPP-4 that only slowly dissociates, explaining a duration of action that persists over the time characterized by high plasma concentrations. The recommended dose in patients with normal renal function is 50 mg twice daily.

**Saxagliptin**

Saxagliptin is (S)-3-Hydroxy-adamantyl-glycine-L-cis-4,5-methano-proline-nitrile (TFA salt) [29,32]. Its structure and mode of interaction with DPP-4 is similar to that of vildagliptin. Saxagliptin is the only DPP-4 inhibitor that is significantly metabolized to an active metabolite (5-hydroxy-saxagliptin), which contributes to the inhibition of DPP-4. Saxagliptin is metabolized by cytochrome P450 (isoenzymes 3A4 and 3A5). Potential interactions with inducers of CYP 450 or other drugs also metabolized by CYP 450 need to be taken into account. In the absence of renal functional impairment or potential interactions, saxagliptin is administered at 5 mg once daily.

**Linagliptin**

Linagliptin is (R)-8-(3-Amino-piperidin-1-yl)-7-but-2-ynyl-3-methyl-1-(4-methyl-quinazolin-2-ylmethyl)-3,7-dihydro-purine-2,6-dione, and has a xanthine-based structure [33]. It is a competitive inhibitor of DPP-4 with a long half-life. One special feature of linagliptin is the absence of any significant renal elimination. Therefore, linagliptin can be administered to all patients, irrespective of renal function, at a dose of 5 mg once daily. Other important pharmacologic properties are shown in Table 48.1. Linagliptin is the only DPP-4 inhibitor with appreciable protein binding in plasma.

**Alogliptin**

Alogliptin is 2-[(6-(3R)-3-Amino-1-piperidinyl)-3-methyl-2,4-dioxo-3,4-dihydro-1(2H)-pyrimidinyl][methyl]benzonitrile [34]. Structurally, it is a modified pyrimidinedione. It has a long half-life and can, in the absence of renal functional impairment, be administered once daily at a dose of 25 mg.

**Efficacy as glucose-lowering agents (reduction in HbA1c)**

**Monotherapy**

All approved DPP-4 inhibitors lower HbA1c when studied in drug-naive patients or at least when used in monotherapy (i.e., after washing out previous antidiabetic drug treatment in some of the patients). The approximate effectiveness is an HbA1c reduction by 0.6–0.9% relative to that seen with placebo treatment in studies lasting approximately 26 weeks. Accordingly, DPP-4 inhibitors can be used as initial drug therapy; however, only in patients who do not tolerate or have contraindications for the use of metformin. Treatment with DPP-4 inhibitors under these conditions is associated with a low risk for hypoglycemic episodes (comparable to placebo treatment) and there is little, if any change in body weight.

**Add-on to oral antidiabetic agents**

All approved DPP-4 inhibitors have been studied as add-on to previously existing treatment with oral antidiabetic agents like metformin, sulfonylureas, thiazolidinediones, and combinations
Table 48.1 Pharmacologic characterization of inhibitors of dipeptidyl peptidase-4 (DPP-4)

<table>
<thead>
<tr>
<th>Compound: Parameter</th>
<th>Sitagliptin</th>
<th>Vildagliptin</th>
<th>Saxagliptin</th>
<th>Linagliptin</th>
<th>Alogliptin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemical structure</td>
<td>β-Amino acid-based</td>
<td>Cyano-pyrrolidide</td>
<td>Cyano-pyrrolidide</td>
<td>Xanthin-based</td>
<td>Modified pyrimidinedione</td>
</tr>
<tr>
<td>DPP-4 inhibition (Type)</td>
<td>Competitive</td>
<td>Competitive</td>
<td>Competitive</td>
<td>Competitive</td>
<td>Competitive</td>
</tr>
<tr>
<td>Enzyme–substrate interaction (Type of bond)</td>
<td>Non-covalent</td>
<td>Covalent</td>
<td>Covalent</td>
<td>Non-covalent</td>
<td>Non-covalent</td>
</tr>
<tr>
<td>Association/dissociation (Velocity)</td>
<td>Rapid</td>
<td>Slow</td>
<td>Slow</td>
<td>Rapid</td>
<td>Rapid</td>
</tr>
<tr>
<td>Pharmacokinetic characteristics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1/2 [h]</td>
<td>8–24</td>
<td>1.5–4.5</td>
<td>2–4 (3–7&lt;sup&gt;a&lt;/sup&gt;)</td>
<td>10–40</td>
<td>12–21</td>
</tr>
<tr>
<td>Dosing frequency Times per day</td>
<td>1</td>
<td>2&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Metabolism (if relevant)</td>
<td>No</td>
<td>Hydrolysis</td>
<td>Active metabolite</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Interaction with CYP 450&lt;sup&gt;c&lt;/sup&gt;</td>
<td>No</td>
<td>No</td>
<td>CYP3A4/5&lt;sup&gt;d&lt;/sup&gt;</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Plasma protein binding</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Predominant route of elimination</td>
<td>Renal</td>
<td>Renal</td>
<td>Renal</td>
<td>Biliary</td>
<td>Renal</td>
</tr>
<tr>
<td>Recommended dose&lt;sup&gt;e&lt;/sup&gt; [mg per day]</td>
<td>&gt;50 mL min&lt;sup&gt;−1&lt;/sup&gt;</td>
<td>100 q.d.</td>
<td>50 b.i.d.</td>
<td>5 q.d.</td>
<td>5 q.d.</td>
</tr>
<tr>
<td>Normal renal function, eGFR</td>
<td>&gt;30–50 mL min&lt;sup&gt;−1&lt;/sup&gt;</td>
<td>50 q.d.</td>
<td>2.5 q.d.</td>
<td>5 q.d.</td>
<td>12.5 q.d.</td>
</tr>
<tr>
<td>Chronic kidney disease, eGFR</td>
<td>&gt;15–30 mL min&lt;sup&gt;−1&lt;/sup&gt;</td>
<td>25</td>
<td>2.5 q.d.</td>
<td>5 q.d.</td>
<td>6.25 q.d.</td>
</tr>
<tr>
<td>&lt;i&gt;eGFR&lt;/i&gt; ≤15 mL min&lt;sup&gt;−1&lt;/sup&gt;, dialysis</td>
<td>25</td>
<td>Not rec.</td>
<td>Not rec.</td>
<td>5</td>
<td>No information available</td>
</tr>
<tr>
<td>Dose reduction strategy</td>
<td>Steps</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>Not nec.</td>
</tr>
<tr>
<td>Specificity&lt;sup&gt;e&lt;/sup&gt; (based on fold selectivity over related peptides)</td>
<td>Overall specificity</td>
<td>High</td>
<td>Moderate</td>
<td>Moderate</td>
<td>Moderate</td>
</tr>
</tbody>
</table>

Not rec., not recommended; Not nec., not necessary

<sup>a</sup>Active metabolite.

<sup>b</sup>50 mg b.i.d. in patients with normal renal function.

<sup>c</sup>Cytochrome P450.

<sup>d</sup>Dose reduction recommended when used in combination with inhibitors of CYP3A4/5, such as ketoconazole.

<sup>e</sup>Versus related proline-specific peptidases.

Source: Adapted from Scheen et al. 2010 [29]; Deacon et al. 2011 [30].

of such agents, if the previous treatment alone did not provide glycemic control with HbA1c values in the target range. If added to metformin, DPP-4 inhibitors typically reduce HbA1c by 0.6–1.0 % relative to placebo treatment (Figure 48.3). Again, this combination is associated with a low risk for hypoglycemic episodes, and most studies have not described major weight changes. Several clinical trials comparing DPP-4 inhibitors with sulfonylureas as add-on treatment to pre-existing metformin treatment describe a similar effectiveness regarding the ability to lower HbA1c, when the study duration was between 6 months and a year. Earlier in the course of treatment, a greater effectiveness is expected for sulfonylureas, but after a treatment duration lasting more than a year, the low “durability” of sulfonylurea-induced treatment effects improved the relative effectiveness of DPP-4 inhibitors (Figure 48.4). Typically, adding a sulfonylurea, will substantially increase the risk for episodes (with approximately one third of the patients reporting such episodes) and will induce body weight gain in the order of 1–2.5 kg.

Adding DPP-4 inhibitors to pre-existing treatment with sulfonylureas (alone or in combination with metformin) lowers HbA1c to a similar degree as seen when adding DPP-4
Figure 48.3 Essential results from clinical trials with different inhibitors of dipeptidyl peptidase-4 (DPP-4) compared to placebo treatment, both administered on a background of metformin treatment. Effect on HbA1c (a), fasting plasma glucose (b), body weight (all Δ versus baseline) and proportion of patients reporting at least one episode of hypoglycemia (d). Trials can be identified by first author and year of publication. Baseline HbA1c (a), trial duration (b), baseline BMI (body mass index) (c), and dose as well as the typical interval between two administrations (d) are also indicated. For references, see [53–60].

Figure 48.4 Essential results from clinical trials with different inhibitors of dipeptidyl peptidase-4 (DPP-4) compared to sulfonylurea treatment, both administered on a background of metformin treatment. Effect on HbA1c (a), fasting plasma glucose (b), body weight (all Δ versus baseline) and proportion of patients reporting at least one episode of hypoglycemia (d). Trials can be identified by first author and year of publication. Baseline HbA1c (a), trial duration (b), baseline BMI (body mass index) (c), and dose as well as the typical interval between two administrations (d) are also indicated. For references, see [44,61–65].
inhibitors to metformin. However, this combination is not characterized by the low risk of hypoglycemia that is typical for DPP-4 inhibitor monotherapy or a combination with metformin (or thiazolidinediones). Since sulfonylurea treatment alone can induce episodes of hypoglycemia, adding a DPP-4 inhibitor (and further lowering the level of glycemia and HbA1c concentrations) may elevate the risk for hypoglycemic episodes. Furthermore, in some studies this combination was associated with some (minor) weight gain.

The question arises whether, in terms of effectiveness, there are clinically relevant differences between different DPP-4 inhibitor molecules in the doses that are recommended for antidiabetic treatment (Table 48.1). All approved DPP-4 inhibitors inhibit serum DPP-4 proteolytic activity by 85% or more; in those in which this has been studied, this is associated with doubling or at most tripling GLP-1 responses after oral glucose or meals. Roughly judging from reported studies based on the ability to reduce HbA1c below baseline, no major differences in clinical effectiveness appear to exist between the different agents.

Add-on to insulin
All DPP-4 inhibitors lower HbA1c when used as add-on to insulin therapy (various regimens) and are approved for this combination [35,36]. However, the clinical trials were not designed to prove that such combinations have the potential to achieve excellent glycemic control. For example, no comparison against the preferred combination insulin plus metformin is available, nor a study proving benefits when DPP-4 inhibitors are added to insulin plus metformin. We recommend using DPP-4 inhibitors as a second choice when metformin is contraindicated or not tolerated. An interesting observation suggests that vildagliptin, when added to insulin, lowers HbA1c and the risk for hypoglycemia at the same time, possibly mediated through an improvement in α-cell responsiveness [36,37].

Comparison to GLP-1 receptor agonists
It is of interest to directly compare therapeutic effects of using GLP-1 receptor agonists (as injectable peptides) and DPP-4 inhibitors (as small molecules that can be used as oral glucose-lowering agents), since both classes of medications are incretin-based, that is, they make use of the therapeutic properties of the parent compound GLP-1 [2]. Exenatide, liraglutide [24], and exenatide once weekly [38] have been compared to sitagliptin in head-to-head comparative clinical trials in type 2 diabetic patients not sufficiently controlled on pre-existing oral antidiabetic treatment. Regarding the control of glycemia, all incretin mimetics were superior to the DPP-4 inhibitor sitagliptin, by approximately 0.3–0.5% in terms of HbA1c reduction compared to baseline values. In addition, treatment with GLP-1 receptor agonists was associated with substantial reductions in body weight (2–3 kg), whereas only minor weight reduction (<1 kg) was observed with DPP-4 inhibitor treatment. Further, the risk for hypoglycemic episodes was not markedly different for an add-on of incretin mimetics and DPP-4 inhibitors. Nausea, vomiting, diarrhea and other gastrointestinal adverse events typically occurred with incretin mimetic, but not with sitagliptin treatment. Thus, there are substantial and typical differences in the results of treating with either GLP-1 receptor agonists or DPP-4 inhibitors. These differences are probably caused by the different levels of GLP-1 (or of the respective incretin mimetic) reached by inhibiting DPP-4 activity (a moderate rise within the “physiologic” concentration range, mostly affecting the gut, mesenteric and portal vein concentrations of GLP-1) or by administering larger doses of incretin mimetics subcutaneously (leading to circulating concentrations in the pharmacologic range, approximately 5–10-fold higher than typical postprandial physiologic levels).

Inability to provoke hypoglycemia
DPP-4 inhibitors exert their antidiabetic effects mainly through augmenting GLP-1 concentrations in response to nutrient intake, with some effect also in the basal state. GLP-1 or GLP-1 receptor agonists, even at high doses, do not themselves provoke hypoglycemic episodes. The biologic explanation is that they do not have the ability to close ATP-dependent potassium channels, which would lead to the depolarization of β cells and the subsequent release of insulin. GLP-1 increases intracellular cAMP levels, which modulate (amplify) insulin secretory responses initiated by hyperglycemia or other “primary” stimuli of insulin secretion. In line with these cell-biologic considerations, in clinical trials DPP-4 inhibitor treatment did not raise the risk for hypoglycemic episodes when used in monotherapy or on top of metformin (and/or thiazolidinediones), for example agents which themselves do not have the potential to provoke hypoglycemic episodes [2]. Rather, there appears to be a potential that endocrine pancreatic α-cell responsiveness to hypoglycemia may be improved by DPP-4 inhibitors. Whether this applies to vildagliptin only, or to other DPP-4 inhibitors as well, is unknown. If confirmed, such findings could provide a rationale to use DPP-4 inhibitors as an adjunct to insulin treatment in patients with T1DM [37] or other forms of diabetes mellitus that require insulin treatment and are associated with a risk for hypoglycemic episodes.

Nevertheless, DPP-4 inhibitors do not prevent hypoglycemic episodes provoked by sulfonylureas. Thus, in combination with sulfonylureas, DPP-4 inhibitors lose one of their typical advantages. Therefore, before using DPP-4 inhibitors together with sulfonylureas, one should carefully weigh potential advantages of such a combination against this obvious disadvantage of allowing for hypoglycemic episodes.

Weight neutrality
In most cases, treatment with DPP-4 inhibitors will not lead to major body weight changes (Figures 48.3 and 48.4). This is in contrast to GLP-1 receptor agonists, which usually induce some weight loss (Figures 48.1 and 48.2), and other classes of glucose-lowering medications like sulfonylureas, meglitnides, thiazolidinediones, and insulin, which have the potential to
induce some weight gain. GLP-1 has the potential to reduce appetite and food intake, and incretin mimetics typically lead to body weight reduction, however DPP-4 inhibitors are “weight neutral” because (a) the degree by which GLP-1 concentrations are raised is too small to exert a significant weight loss; (b) DPP-4 inhibition reduces the level of glycemia and associated glucosuria, thus without appetite reduction, some weight gain would be expected; (c) DPP-4 not only acts on GLP-1 (which is degraded and inactivated), but also on another gut peptide from the L cell, peptide tyrosine tyrosine (PYY), which is activated through DPP-4 action. DPP-4 inhibition potentially takes away weight-lowering effects mediated through GLP-1.

**Adverse events**

Originally, the long list of peptides with biologic activity sharing the structural features required to be a substrate of DPP-4 (i.e., with an alanine or proline residue in the penultimate amino-terminal amino acid position), gave rise to the expectation that any alteration in biologic activity in any of these putative DPP-4 substrates could be a basis for adverse events. However, no specific side effects could be predicted based on such biochemical knowledge. In contrast to expectations, adverse events reported from clinical studies involving DPP-4 inhibitors have not been shown to deviate with any frequency from those typical for placebo (in double-blind studies) or other glucose-lowering therapies with well-accepted adverse-event profiles [39]. Nor has withdrawal of DPP-4 inhibitor treatment due to drug-induced adverse events been reported with any of the agents with any noticeable frequency. Initial safety summaries reported a slightly greater occurrence of upper respiratory tract infections with DPP-4 inhibitor treatment, putatively related to a role of DPP-4 (namely, lymphocyte marker CD 26) in the immune system. More recent summaries encompassing a greater number of patients studied no longer support this finding. Certainly, DPP-4 inhibitors do not cause gastrointestinal adverse events in the same way that incretin mimetics can. There are rare reports of severe skin-related adverse events (Stephens–Johnson syndrome etc.), but a causal role regarding any one of the DPP-4 inhibitors has not been established.

**Concerns and expectations regarding incretin-based therapies**

To date, numerous clinical trials studying the efficacy and safety of incretin-based glucose-lowering medications have been reported, both in the classes of GLP-1 receptor agonists and DPP-4 inhibitors. While these studies provide a database for judging clinical efficacy and safety, longer term studies with clinical outcomes providing evidence that diabetic complications can be prevented or delayed are still lacking. Such studies are under way. Most of them were designed primarily to prove the cardiovascular safety of the drug in question, with a chance to document potential superiority versus a specific comparator or—in most cases—standard treatment. Outcome trials for the incretin-based glucose-lowering medications will report from 2015 on. Meanwhile, it is worth looking at more preliminary sources of evidence.

**Cardiovascular risk**

GLP-1 receptor agonists lower glucose concentrations, body weight, and systolic blood pressure. In addition, some nonclassical cardiovascular risk factors have been found to be reduced. These include markers of insulin sensitivity, of low-grade inflammatory responses (like high-sensitive C-reactive protein), and lower postprandial excursions of glucose and triglyceride concentrations (potentially related to a deceleration in gastric emptying). In addition, specific interactions of GLP-1 and of exenatide and lixisenatide with the cardiovascular system have been described. This relates to endothelium-dependent vasodilatation and to the prevention of large necroses after induced acute myocardial infarction, with salvage of ischemic myocardium at risk induced by GLP-1 receptor stimulation [40]. Furthermore, in animal models of congestive heart failure, GLP-1 can improve pump function along with improved uptake of oxygen and glucose into the myocardium. Clinical pilot studies have supported the idea of beneficial influences of GLP-1 receptor stimulation in the short term [40]. The mechanisms are not quite clear. GLP-1 receptors (of the well-described “pancreatic” type) are expressed in the cardiovascular system, but in pharmacologic studies, cardiovascular effects have been described not only for GLP-1 of the intact, biologically active sequence (GLP-1 (7-36 amide) or (9-37)), but also for the DPP-4 degradation products (GLP-1 (9-36 amide) or (9-37)), which are inactive at the pancreatic GLP-1 receptor. Conversely, exendin-4, which is a high-affinity ligand at the pancreatic type GLP-1 receptor, was sometimes found to be less effective than GLP-1. Taken together, these findings suggest that there may be a second GLP-1 receptor, specifically expressed in the cardiovascular system, with a substrate specificity different from the well-characterized GLP-1 receptor that mediates the insulino-tropic activity of GLP-1 at the level of the β cell [40].

While clinical trials with cardiovascular events as their main endpoints are under way with most of the approved incretin-based medications, preliminary analyses based on adverse events reported from phase 3 trials indicate a trend for reduced numbers of acute myocardial infarctions, stroke, and cardiovascular death [41]. Therefore, it is reasonable to hope that these trials will confirm the cardiovascular safety of GLP-1 receptor agonists and DPP-4 inhibitors, with a potential to detect a benefit, either based on the immediate consequences of improved glycemic control, the effects on body weight and cardiovascular risk factors associated with such changes, or based on additional, specific interactions with the cardiovascular system.

**Acute pancreatitis**

Acute pancreatitis has been diagnosed in patients treated with GLP-1 receptor agonists or DPP-4 inhibitors. The question is
whether there is any causal relationship, leading to an increased risk for acute (or chronic) pancreatitis induced by such treatment. Animal experiments have described histologic changes compatible with acute or chronic pancreatitis after treatment with sitagliptin and exenatide, but with different experimental design, no such changes or even anti-inflammatory changes have been documented with liraglutide treatment. Epidemiologically, several analyses of claims databases have uniformly described the absence of an elevated risk for pancreatitis [41]. The notable exception is an analysis of the FDA adverse events reporting system that found a several-fold elevation in the risk of pancreatitis as well as pancreatic cancer [42]. The circumstances of this analysis (i.e., the timing of collecting data) suggest the influence of a number of different potential reporting biases. In all, the question of incretin-based medications promoting pancreatitis must be regarded as an open one, however, with the bulk of available data not supporting a major risk [41]. Acute pancreatitis usually is diagnosed when at least two out of three criteria are fulfilled: (a) typical symptoms including severe abdominal pain, usually leading to presentation at an emergency department; (b) elevated lipase or amylase activity in plasma; and (c) findings supporting the diagnosis using imaging procedures. Recent findings from clinical trials with incretin-based medications have shown that patients with T2DM often (approximately 20%) display elevated pancreatic enzyme levels at screening examinations, even when they did not report symptoms suggesting pancreatitis. In a study employing liraglutide treatment, pancreatic enzyme levels slightly rose after exposure to the drug. While this indicates some interaction of GLP-1 receptor stimulation with the exocrine pancreas, it does not necessarily indicate the initiation of inflammatory changes. The nature and consequences of this interaction need to be better defined. However, it appears possible that acute pancreatitis has been over-diagnosed in patients with elevated enzymes, but without appropriate symptoms (and ensuing work-up) supporting the diagnosis of acute pancreatitis, especially in patients treated with drugs that typically cause gastrointestinal adverse events. In conclusion, it appears wise to suggest caution when using GLP-1 receptor agonists in patients with a history of acute pancreatitis, and to discontinue treatment if suspicion or diagnosis of acute pancreatitis arise.

**Pancreatic carcinoma**

Some animal studies have described histologic changes compatible with chronic pancreatitis and an increased pancreatic ductal proliferation rate after treatment with exenatide or sitagliptin. Sometimes, this was associated with more pancreatic duct glands (PDGs: outpouches from major pancreatic ducts) or low-grade pancreatic intra-epithelial neoplastic lesions (PanINs), which are thought to represent preneoplastic lesions associated with the transition from chronic pancreatitis to pancreatic carcinoma. On the other hand, the bulk of animal toxicology data from all development programs of incretin-based glucose-lowering medications has not reported any such changes, and specific studies with long-term, high-dose exposure of animals to incretin-based drugs have not confirmed a uniform influence of incretin-based medications to support the development of such histologic changes. Case reports of pancreatic cancer developing after initiating incretin-based treatment in patients with a previously normal pancreatic morphology have not been reported, and the natural history of the development of pancreatic cancer is thought to be a slow process, with little chance to alter its course within a few years (the longest exposure possible to date). The more surprising was the finding, again from the analysis of the FDA adverse events reporting system, that exenatide and sitagliptin treatment were associated with an elevated risk for the diagnosis of pancreatic cancer [42]. Again, better surveillance of such events with incretin-based medications is necessary to define or rule-out a potential risk in the long term [41].

**C-cells, medullary thyroid carcinoma**

In rodents, there is a high prevalence of C-cell abnormalities including hyperplasia, adenomas, and spontaneous C-cell (medullary thyroid) carcinomas. Liraglutide treatment increases the prevalence of C-cell disease, including medullary carcinomas, especially in male rats. This finding gave rise to the question whether there could be a risk of increasing the incidence of this rare endocrine tumor in patients treated with GLP-1 receptor agonists. Again, a higher risk for thyroid carcinoma (irrespective of the histologic type) was suggested by the analysis of the FDA adverse events reporting system [42]. In rodent C-cells, GLP-1 receptor expression is prominent, their stimulation (by any ligand) will raise cAMP and lead to secretion of calcitonin, and probably induces proliferative changes [43]. In human C-cells, cAMP levels and calcitonin are not affected, so it is questionable whether any proliferative responses can be expected [43]. Whether or not follicular cells and malignant cells from different varieties of thyroid carcinomas express GLP-1 receptors is unknown. Controversial findings have been reported. Cases of thyroid carcinoma diagnosed after initiation of GLP-1 receptor agonist treatment in patients with previously normal thyroid morphology have not been reported. In all, at present the possibility of inducing a higher risk for medullary (or other histologic types of) thyroid carcinoma can neither be totally refuted nor confirmed. Better surveillance will hopefully clarify this point in the long term. Based on the low incidence of especially the medullary variety of thyroid carcinomas the risk will most likely be low. Patients with a genetic trait predisposing to medullary thyroid carcinomas (hereditary medullary thyroid carcinoma or multiple endocrine neoplasia type 2) should not be treated with GLP-1 receptor agonists. There are no critical findings regarding DPP-4 inhibitors with respect to thyroid carcinomas.

**Clinical use of incretin-based medications**

While clinical trials in patients who had been drug-naïve or washed-out from previous treatment with oral antidiabetic
agents have shown the clinical effectiveness of both GLP-1 receptor agonists and DPP-4 inhibitors, they usually are employed in this patient population only when metformin is contraindicated or not tolerated. In most patients, DPP-4 inhibitors or GLP-1 receptor agonists are used as a second- or third-line glucose-lowering agent, added to previously used metformin treatment, when glycemic control is no longer adequate with monotherapy. In principle, they compete, at this stage of treating T2DM, with sulfonylureas or meglitinides, with thiazolidinediones, α-glucosidase inhibitors, or insulin. DPP-4 inhibitors, which are administered as tablets for oral ingestion, tend to be used at earlier stages, when injectable therapy does not need to be considered to achieve individual glycemic targets. GLP-1 receptor agonists, as injectable treatment, will typically be considered at later stages, when otherwise insulin would have to be considered as an alternative treatment.

**DPP-4 inhibitors versus sulfonylureas**

As add-on to metformin treatment, sulfonylureas had been popular; however, their use was associated with a risk for hypoglycemia (sometimes requiring professional help or even hospitalization, and in rare cases leading to death) and with weight gain. In addition, epidemiologic data have suggested some risk for more cardiovascular events and increased mortality, although this has not been confirmed in prospective studies, especially the UKPDS. As a result of these potential adverse effects, the use of sulfonylureas has been declining over the past years. Rather, DPP-4 inhibitors have been increasingly used. Head-to-head comparisons (e.g. Figure 48.4) have shown that the ability of a variety of DPP-4 inhibitors to control glycemia is comparable to the effectiveness displayed by sulfonylureas (Figure 48.4). Most comparisons observed a better reduction in fasting glucose concentrations by sulfonylureas, indicating a more pronounced postprandial action of DPP-4 inhibitors. DPP-4 inhibitors induce some minor weight reduction, while sulfonylureas uniformly raise body weight by approximately 1–2 kg. Up to one third of the patients treated with sulfonylureas experience at least one episode of hypoglycemia over a period of 26–52 weeks, possibly with differences between different members of the sulfonylurea class. Of particular interest may be the fact that in a direct head-to-head comparison of the sulfonylurea glimepiride with the DPP-4 inhibitor linaagliptin there was a significant difference in cardiovascular outcomes in favor of linaagliptin treatment, albeit at a relatively low total number of events [44]. A similar trend had been noted for the comparison of other sulfonylureas with a variety of DPP-4 inhibitors previously. Thus, for patients who need to avoid hypoglycemia or who intend to reduce body weight, DPP-4 inhibitors may be preferable.

**GLP-1 receptor agonists versus insulin**

As injectable therapies, insulin and GLP-1 receptor treatment are often considered when a combination of oral antidiabetic agents is no longer able to control glycemia within individually agreed treatment targets for HbA1c and plasma glucose. Theoretically, insulin treatment can achieve any reduction in plasma glucose concentrations that is required, even in severely hyperglycemic patients, depending on finding and employing the appropriate dose(s). GLP-1 receptor agonists, which are used at standard doses that are usually not varied with the effects achieved, have whatever clinical effectiveness is associated with the use of the particular dosage. Thus, the preoccupation was that comparing GLP-1 receptor agonist treatment with insulin treatment would be a bold comparison, with a great likelihood that insulin therapy would be more effective. In several clinical trials comparing exenatide (twice daily), liraglutide, or exenatide once weekly with either insulin glargine or premixed insulin (Figure 48.4), the effectiveness of using a GLP-1 receptor agonist was at least equal to that of insulin treatment. In some, but not all of these trials, insulin was better in controlling fasting glucose, indicating a more pronounced postprandial mode of action (deceleration of gastric emptying etc.) in the case of GLP-1 receptor agonists. It had to be discussed whether the dosage of insulin used was adequate and allowed a full exploitation of the therapeutic potential inherent in insulin-based treatment regimens. Most likely, insulin titration in such clinical studies could potentially be optimized, but was very similar to the way insulin treatment is administered in clinical practice. Insulin treatment, even intensified regimens with multiple daily injections, do not guarantee goal achievement. Experience or fear of weight gain or of hypoglycemic episodes may limit efforts to stringently increase insulin doses as determined by a treat-to-target algorithm. Along these lines, GLP-1 receptor agonist treatment may be particularly better in reducing HbA1c in the patient most difficult to treat, such as those with high baseline HbA1c. Based on this experience, it may no longer be justified to assume that insulin-based treatment regimens are the ultimate escalation step in the treatment of T2DM patients, for whom other approaches have failed. Of course, there is the possibility to combine long-acting insulin with GLP-1 receptor agonists [19,20], a regimen that may successfully compete with intensified insulin regimens as the most effective treatment regimen for those not achieving their goals with oral antidiabetic agents and long-acting insulin alone.

**Recommendations for special populations**

Subgroups of patients with special characteristics, such as higher age, impaired renal function, impaired kidney function, or gastrointestinal diseases, which could potentially interfere with the mode of action of a therapy that depends on the interaction of a gastrointestinal hormone (GLP-1) with its receptor, need special consideration regarding the use of incretin-based glucose-lowering medications.

**Renal functional impairment**

GLP-1, like other gastrointestinal hormones (including the partner incretin hormone GIP), is eliminated renally. Patients with significant functional impairment of their kidneys (i.e., a
Liver diseases that need to be considered are diabetic fatty liver (steatosis or steatohepatitis), which are highly prevalent in an obese, type 2 diabetic population, and advanced liver disease (fibrosis, cirrhosis), which may be associated with impaired liver function as far as synthesis/secretion of liver-derived proteins including albumin, blood clotting factors, etc., and detoxification/elimination function are concerned.

Both obesity, caloric excess, and loss of glycemic control predispose to fatty liver. Besides thiazolidinediones, which specifically remove triglycerides from the hepatic parenchyma and promote storage in subcutaneous adipose tissue depots, GLP-1 receptor agonists are a second class of glucose-lowering agents that may be associated with a specific benefit in patients with fatty liver disease. Supporting findings are from animal experiments or preliminary clinical evidence. The main effect may be a consequence of incretin-mimetic-induced weight loss, but more specific interactions with hepatic triglyceride metabolism, inflammation, etc. appear possible. Among the DPP-4 inhibitors, vildagliptin is hepatically metabolized and sometimes has increased liver enzymes, leading to the recommendation of monitoring transaminases. Thus, vildagliptin is not the preferred treatment in the face of liver disease.

GLP-1 receptor agonists may delay gastric emptying. Their use should be avoided in patients with clinically significant retardations in gastric emptying, that is, with a diagnosis of severe autonomous neuropathy of the gastrointestinal tract including symptomatic diabetic gastroparesis. This, however, is rarely diagnosed in patients with T2DM, even after long duration of the disease. A deceleration of gastric emptying may accentuate regurgitation of gastric content in patients prone to gastroesophageal reflux disease. Patients with constipation due to intestinal autonomic neuropathy may be at risk for worsening symptoms when treated with incretin mimetics. These considerations do not apply to the clinical use of DPP-4 inhibitors.
Association and the European Association for the Study of Diabetes in 2009 mentioned GLP-1 receptor agonists, however, among “less well-validated therapies” for T2DM [46]. An official statement of the American Association of Clinical Endocrinologists from 2009 gives specific recommendations for the use of both classes of incretin-based antidiabetic medications, with an emphasis on dual and triple therapy including such agents [47]. The British National Institute for Health and Care Excellence (NICE) clinical guideline 87 gives detailed recommendations for employing GLP-1 receptor agonists and DPP-4 inhibitors,
for example suggesting entry criteria for such treatment (like a minimum degree of obesity in the case of incretin mimetics) and recommending continued treatment with such agents only when clinical effects have been documented individually (http://www.nice.org.uk/nicemedia/pdf/CG87QuickRefGuide.pdf). The latest position statement by the American Diabetes Association and the European Association for the Study of Diabetes issued in 2012 is based on the principle of individualized approaches offering medications that best help attain individual treatment targets. Since these often include the avoidance of hypoglycemic episodes and weight gain, such objectives may speak in favor of using incretin-based medications [48].

**Future developments**

In the coming years, we can expect important new developments in the treatment of T2DM with incretin-based drugs, both regarding new agents in both classes and more results of clinical trials with relevant clinical endpoints, including cardiovascular outcomes. The latter information will certainly help to define the value of incretin-based medications for the prevention of diabetic complications. Such studies will also help understand the open questions regarding pancreatitis and pancreatic carcinoma, as well as (medullary) thyroid cancer. Regarding novel compounds within the classes of GLP-1 receptor agonists and DPP-4 inhibitors, a substantial number of incretin mimetics are under development, among them albiglutide, lixisenatide, dulaglutide, semaglutide. They are mainly designed for less frequent administration, for example once-weekly injections, which, if hoped will improve adherence to treatment. Whether it will be possible to improve the relationship between effectiveness and side effects is an open question. Given the differences observed between different compounds, there seems to be space for further improvements. Within the class of DPP-4 inhibitors, this appears more difficult, since all available compounds have similar glucose-lowering potency and are virtually free from drug-induced adverse events. There is an opportunity for longer inter-administration intervals in the class of DPP-4 inhibitors as well. Deeper understanding of the mechanisms of action, however, may present clues for further improvements regarding effectiveness as well.

**References**

17. Diamant M, Van Gaal L, Stranks S, et al.: Once weekly exenatide compared with insulin glargine titrated to target in patients with


CHAPTER 49

Insulin pumps

Dace L. Trence and Irl B. Hirsch
University of Washington School of Medicine, University of Washington Medical Center-Roosevelt, Seattle, WA, USA

Key points

• Insulin pump technology has advanced to now include computer-assisted insulin dosing, incorporating glucose data, carbohydrate counting, and insulin tracking.
• Continuous glucose sensors added to insulin pump technology have further advanced self-management of diabetes mellitus.
• Optimal patient use of this technology requires support in learning how to use these tools, and in the case of continuous glucose sensors, requires ongoing patient attention and use of data obtained.
• Future advances in the integration of insulin delivery in response to glucose sensing are evolving with attention to low-glucose sensing and automatic pump insulin delivery as a first step toward a fully automated artificial pancreas model.
• Clinicians should be aware and ideally comfortable with evolving technology to facilitate the ability to present, as well as support, these tools in directing and supporting the self-management of diabetes of their patients.

History

Insulin pump or continuous subcutaneous insulin infusion (CSII) technology was in its infancy in the 1960s when Dr. Arnold Kadish in Los Angeles, CA first conceptualized a closed loop system which provided the components of blood glucose sensing coupled with insulin infusion [1]. Designed to be worn on the back to facilitate intravenous access, it was certainly portable but was also very large and cumbersome which proved to be a major drawback. This insulin pump prototype was comprised of a glucose autoanalyzer to measure blood glucose, coupled to an on/off switch operating a pump that delivered insulin. When the autoanalyzer detected glucose levels outside the normal range, insulin was automatically turned off. In the next decade, a new device called the Biostater became available and touted computer-controlled glucose sensing and insulin delivery system. However, it also fell to the wayside due to its unwieldy size and lack of portability.

The Biostater consisted of multiple components: a pump that withdrew blood for analyses, a glucose analyzer for continuous glucose determinations, computerized algorithms for calculating insulin or dextrose amounts to be infused based on glucose blood levels, a computer-driven infusion pump for insulin or dextrose delivery, and a print out of blood readings based on per minute determinations [1]. Primarily, this device was used in research, although it was also used in obstetrical arenas.

The next iteration of the insulin pump was designed to be worn as a shoulder bag in an attempt to overcome the portability issues of previous designs. Slama et al. [2] reported on this insulin delivery system that provided insulin over 1–5 days, a basal insulin at a constant rate, with prandial insulin boluses at a set 15-fold higher rate. Unfortunately, progress with this model was hampered by insulin delivery being intravenous and the consequent risks of infection, thrombosis, and phlebitis. Further advancement of a portable design was accomplished at the National Institute for Medical Research in England. The Mill Hill Infuser, first developed for research entailing infusing hormones into animals, could deliver insulin in a dual rate process: a basal rate and a prandial rate that was eightfold higher. Delivery was available at the push of a button with a battery-driven syringe pump of a dramatically reduced size and weight of only 5.6 ounces—about a third of a pound [3].

Technology continued to progress with the first commercial pump sold on the market in 1978 under the name AutoSyringe, commonly referred to as the “Big Blue Brick” due to its appearance. Early sales surpassed expectations at 600 pumps per month, sparking interest in commercial pump development by many device companies in the early 1980s [4]. Numerous setbacks occurred at first due to problems with safety of insulin delivery, limited safety alarms, little flexibility in insulin dosing, needles that were metal and dislodged frequently, and batteries that required frequent recharging [1]. Some models even required a screwdriver to adjust insulin doses (http://www.medscape.org/viewarticle/460365_2 (accessed August 27, 2012)). Not surprisingly, complications from lack of glucose
control, such as diabetic ketoacidosis, hyperglycemia, and infection at the insertion site were common [5]. Bulksiness continued to impair the appeal of insulin pumps as many weighed up to a pound and were the size of a house brick. These early pump prototypes were more likely to be prescribed for those with diabetes that was difficult to manage, not infrequently giving less than expected benefit, in turn dampening patient as well as prescribing clinician enthusiasm for personal pumps.

The 1990s provided a true evolution in technological advances that allowed insulin pumps to become much safer and reliable devices. Improvements to reduce insulin leakage, provide more consistent insulin delivery, and address tubing occlusion, along with more reliable alarms for these potential malfunctions, boosted user safety. Alarms for low battery or low insulin amount in the pump reservoir increased users’ sense of security in depending on insulin pump therapy. Clearly, more reliable pump devices led to a new level of comfort in prescribing these for more people with diabetes, not limiting this technology to those with only the most challenging need to treat diabetes.

Pumps now function even more effectively by storing multiple preprogrammed basal insulin rates in their memory. Bolus doses can be determined by entering the carbohydrate content of an upcoming meal, as well as a glucose target range, and a correction factor for not only pre-meal glucose adjustment, but also for glucose correction outside the meal setting. The concept of insulin on board based on the pharmacodynamics of rapid analogue insulin [6] allows for the ability of the pump to calculate a more precise correction insulin dose—typically a more beneficial dose than would be apparent to the pump user, who might be tempted to guess at a correction dose! These so-called “smart pumps” can also deliver insulin boluses to prolong the action of the insulin activity. Extended boluses are beneficial for higher fat and protein meals such as a burger and fries or pizza—where the impact of carbohydrate can be dramatically delayed. The risk of pre-meal hyperglycemia is also reduced as extended boluses can be given with a proportion of the total dose initially given normally. Unfortunately, there is a learning curve with this feature and a lot of trial and error is necessary for maximum efficacy. Nevertheless, the introduction of continuous glucose monitoring has taken some of the guess work out of extended boluses. Another aspect of extended boluses allows for an increase in temporary basal insulin. Situations in which the user is engaged in a prolonged period of snacking, such as a wedding reception, are ideal for this feature as snack foods are typically high in carbohydrates.

Additionally, pump manufacturers have listened to their customers as evidenced by the variety of options now available. Pump encasements come in a rainbow of colors, LED screens and digital displays ensure easier visualization of menu options, and tubeless insulin delivery systems based on a pod containing an insulin reservoir and pump are operated by a remote computerized hand-held device allowing more discretion. With regard to size, pumps are dramatically smaller and resemble a credit card in size. Such developments have device manufacturers appealing to a growing market of demand. The global insulin pump market was valued at $586.4 million in 2009, and is forecast to grow at a compound annual growth rate of 9% from 2009 to 2016. Market growth is attributed in part to rising patient awareness. In the US for the same time frame, the same rate of growth is expected bringing it to $843 million in 2016. The traditional insulin pump segment was valued at $394 million in 2009. And there is an emerging disposable insulin pump market, valued at an additional $58.9 million in 2009. Clearly pump therapy has developing appeal and the US holds the largest share of the global market (http://www.mtbeurope.info/content/ft1008001.htm (accessed September 3, 2012)).

Further enhancing pump technology is the ability to link it to continuous glucose sensing. By combining the ability to look at glucose trends to allow for insulin adjustment and deliver insulin the concept of an “artificial pancreas” may be realized. Current technology is not yet advanced to the point where the pump user can become a “disinterested bystander.” A system that senses low glucose via continuous sensor (a glucose level programmed by the user) resulting in an automatic suspension of insulin delivery by pump for 2 hours, then automatic restart at preprogrammed basal delivery rate, has been released for use in most of the world and at the end of 2013, also in the United States.

Along with reviewing the rewards of constantly evolving technology it is also prudent to consider the drawbacks and identify the future course of insulin pump therapy. Clearly more sophisticated features have improved safety. Notably, pumps boast programmable memory, lockout alarms, lack of bolus over time alarms, maximum bolus delivery locks, and ability to download pump memory to check insulin delivery. Even with numerous safety features, there are no guarantees since users may ignore alarms, forget settings and fail to check them when replacement or refurbished pumps are needed. Additionally, successful pump therapy relies on two key components prior to initiating pump therapy. First of all, accurate pump settings need to be determined by a team working closely with the user [7]. Secondly, although there are clinical situations where pump therapy can be more successful than multiple daily insulin dosing (MDI) injection therapies [7], the ability to optimally improve glycemic control routinely depends on extensive individualized education prior to initiating pump therapy. Typically, the process requires education in insulin pharmacokinetics as well as pharmacodynamics, the concept of lag time between insulin bolus and meal ingestion, the use of correction dose insulin, and carbohydrate counting [7]. The use of a multidisciplinary team for pump therapy cannot be underestimated for best results. Even friends or community members could potentially act as resources, although data to support this is needed. It is both notable and unfortunate that a multidisciplinary team approach is limited in the majority of primary care, and even in many endocrinologists’ clinical practices.
**Efficacy, benefits, data supporting insulin pump therapy**

We are in an age that demands the use of scientifically evidence-based practices for prescribed care modalities. Given the typical $7000 price tag to get started on a pump, even users with an average insurance policy covering 80% find themselves burdened with a large expense. Both users and providers must consider whether the investment is justified by the benefits. For starters, where has pump therapy been shown to have a favorable impact on glycemic control? Are there indications that it helps to prevent the complications of diabetes? And are there specific populations, for which data suggest a stronger benefit, or for whom there is a limiting factor to consider for the potential use of an insulin pump?

Whether the pump user is young or old, studies reveal that age is not a limiting factor. Plotnick [8] followed 95 patients who began insulin pump therapy with a mean age of 12.0 years (range 4–18), 29% being <10 years old, for a median duration of 28 months. A small but significant decrease in HbA1c at 3–6 months after pump start (7.7 vs. 7.5%; \( p = 0.03 \)) was followed by a gradual HbA1c increase. After adjusting for diabetes duration and age, mean HbA1c after pump start was significantly lower than before pump start (7.7 vs. 8.1%; \( p < 0.001 \)). The number of medical complications (diabetic ketoacidosis, emergency department visits) was similar before and after pump start. On the plus side, there tended to be fewer hypoglycemic events after pump start (12 vs. 17; rate ratio 0.46; 95% CI 0.21–1.01).

Sulli [9] reported on 42 patients (mean age 12.2 +/- 3.4 years; range 4.5–17 years; 24 males; mean duration of T1DM 5.1 +/- 3.0 years) followed over 4 years. Mean HbA1c levels prior to pump initiation at 8.9 +/- 1.0%, decreased during the first (8.2 +/- 0.9%; \( p = 0.00 \)), second (8.6 +/- 1.0%; \( p = 0.05 \)), third (8.4 +/- 0.9%; \( p = 0.01 \)), and fourth (8.2 +/- 1%; \( p = 0.00 \)) year of pump (CSII) therapy. Insulin requirements (U kg^{-1} per day) decreased during CSII treatment compared with multiple daily insulin dosing (MDI) treatment. Through the first, second, third, and fourth years of follow-up the number of episodes of severe hypoglycemia (20.0, 20.0, 20.0, and 0 episodes/1000 patient-years, respectively) and diabetic ketoacidosis (0.05, 0.00, 0.03, and 0.00 episodes/patient per year, respectively), events were similar to that in the year preceding CSII therapy (2.0 and 0.03, respectively), suggesting no increase in ketoacidosis episodes on CSII therapy in this young population.

De Bock [10] reported a 6-year retrospective analysis of children with T1DM started on insulin-pump therapy from 2002 to 2008 compared to the MDI user T1DM population matched by age, sex, ethnicity, and duration of diabetes. Transitioning to insulin-pump treatment was associated with an improvement in HbA1c compared with baseline (~0.3% per year; \( p < 0.001 \)) up to 3 years. In contrast, nonpump controls showed a continuing trend to higher HbA1c values (+0.2% per year; \( p < 0.01 \)). The risk of severe hypoglycemia fell after pump start (from 27 to 5 events/100 patient-years) with no change in nonpump controls; the rate of diabetic ketoacidosis was low in both groups, and not different.

Fuld [11] reviewed the literature on children and insulin pump use, and noted the discrepancies between randomized and nonrandomized control study results with nonrandomized studies showing a consistent fall in HbA1c of up to 0.9% with reduced frequency of hypoglycemia in the patients switched from MDI to CSII. This was not consistently seen in the randomized studies. However, Kapellen [12] reported on 1567 German and Austrian children and adolescents with T1DM who were placed on CSII between 1995 and 2005 from MDI, 138 of whom were under 5 years of age. The major indication (42.5%) for switching from MDI to CSII was to reduce severe hypoglycemic events. And indeed the rate of severe hypoglycemic events decreased from 52.1 per 100 patient-years to 24.8 per 100 patient-years (\( p < 0.001 \)) after 1 year of pump therapy. Additionally these lower rates were not associated with a deterioration of glycemic control (HbA1c, 7.6% vs. 7.5%). So although there is inconsistency in HbA1c impact, there is clearly less hypoglycemia, a suggestion of less weight gain, and no increase in ketoacidosis in children on CSII.

In contrast, quality of life improves for parents whose children are started on CSII, as evidenced by multiple studies [13]. The studies also noted that children did not return to MDI after being switched to CSII when the studies were completed, reflecting a preference for pump therapy. Hypoglycemia and related seizures are a major parental fear, which can adversely impact the ability to find appropriate childcare. Interestingly, Weinzimer [14] found that children who received daytime care from paid caregivers, such as nannies and daycare centers, had an even greater improvement in HbA1c levels, from 7.5 +/- 0.9% pre pump to 7.1 +/- 0.8% post pump (\( p = 0.002 \)), provided caregivers had been instructed on CSII and had ready access to parents via telephone or beeper.

Sullivan-Bolyai et al. [15] interviewed parents of children on pump therapy to assess their experiences. Despite initial hesitation about pump therapy, parents reported becoming comfortable between 10 days to 3 months after the start of CSII. In fact, all family members reported feeling more freedom and flexibility in their daily lives. Improved marital relationships were also reported, as well as the benefits that CSII “gave them their children back,” children were “in a better mood” and had increased school concentration. Even in comparing CSII to MDI in preschoolers, specific to psychosocial measures targeting parent and child psychological adjustment, parents reported a significant decrease in diabetes-related worry with CSII [16]. Clearly, the studies are consistent in showing improved quality of life in families with a child on CSII, which led to a position statement in 2006, by the Lawson Wilkins Pediatric Endocrine Society [17] that stated that all children with diabetes, regardless of age, should be “considered potentially eligible” for CSII. They further stressed that families should be made fully aware of
the expectations of CSII treatment and what CSII “can and cannot do.”

In the elderly, there is little evidence supporting that older patients treated with insulin pump therapy have improved clinical outcomes compared to patients treated with MDI [18]. The effectiveness of MDI with currently available insulin analogues used in basal-bolus insulin regimens provides a very effective alternative to pump therapy. Hirsch et al. [19] reported 100 patients with T1DM from all age groups randomized to either MDI using aspart before meals and glargine at bedtime or to CSII for 5 weeks and then crossed patients over to the alternate treatment group for an additional 5 weeks. Although fructosamine levels were significantly lower and glucose exposure profiles improved in the CSII periods of treatment, no differences were noted in the frequency of hypoglycemia. The mean age, was 43 ± 11.1 years with a range of 30–55 years, and there was no stratification of outcomes by specific age. Regardless, it is reasonable to assume that similar outcomes would be noted in the youngest as well as the oldest groups of patients.

A Cochrane Review [20] evaluated 23 studies with 976 participants with T1DM randomized to CSII versus MDI and noted a statistically significant difference in HbA1c favoring CSII (weighted mean difference, −0.3 (95% CI, −0.1 – 0.4)). Also noted were that patients with severe hypoglycemia benefited from CSII and reported improvements on quality of life measures. No significant differences were seen in body weight changes, none severe hypoglycemia events, morbidity, mortality, or costs. However, the conclusion for cost in this review was curious, as it was not based on differences in benefits, and was referred to as not available in any study, but was none the less estimated as the additional cost associated with using CSII over MDI, reported as varying from £1091 (€1208, 11/2009 exchange rate) to £1680 (€1860, 11/2009 exchange rate), per year (including the cost of the pump and consumables). Despite some evidence for benefit with CSII, the Cochrane Review dataset had few patients in older age groups, making it difficult to extrapolate its findings to the elderly.

Only one randomized trial has compared efficacy and quality of life in older adults (mean age, 66 years) randomized to either CSII or MDI. In this study, Herman et al. [21] followed 107 adults with T2DM (mean HbA1c, 8.2%) treated with insulin for 12 months. Patients were not selected on the basis of having failed MDI therapy or having achieved suboptimal results. No statistically significant differences were noted in HbA1c lowering, hypoglycemia frequency, weight gain, treatment satisfaction, or quality of life measures between the MDI and CSII groups. However, Rizvi et al. [22] sought to describe the experience of a subset of older patients (mean age, 66.4 years) with T1DM who were characterized by a mean duration of disease of 33 years, suboptimal glucose control with MDI, and frequent hypoglycemic events. Although this study was limited by a very small cohort, after transition to CSII, the mean HbA1c value in this series of five patients, decreased from 9.16% to 7.6% ($p < 0.0025$), and the adjusted frequency of severe hypoglycemia decreased from 3.2 episodes to 0.4 episodes per year ($p < 0.02$) during an average follow-up of 12.6 months. Given these results, there is reason to suggest a benefit of decreased hypoglycemia in the elderly as was seen in the young.

Additionally, special populations such as the newly diagnosed individual with T1DM have been examined as to their potential benefit from CSII. Ramchandani et al. [23] studied 28 patients with newly diagnosed T1DM, mean age of 12.1 ± 6.2 years initiated on CSII, as early as within 1 day of being diagnosed. All subjects accepted CSII when offered and none elected to discontinue CSII after follow-up periods of up to 3 years—suggesting clear therapy acceptance. Expectedly, HbA1c levels declined from an initial mean of 10.5 ± 2.4% to between 6.5% and 7.4% over the next 18 months, at a mean insulin requirement of 0.33 units kg$^{-1}$ d$^{-1}$ at 3 months, which by 18 months rose to 0.58 units kg$^{-1}$ d$^{-1}$. C-peptide values remained stable during the first 12 months after diagnosis; but there was no significant weight gain for the duration of the study (20.7 kg m$^{-2}$ vs. a peak of 22.1 kg m$^{-2}$ at 12 months; $p = 0.54$). Indeed, early initiation of pump therapy might better preserve β-cell function than MDI. Thrailkill [24] reported a randomized controlled study in 24 patients, 8–18 years old, with newly diagnosed T1DM, evaluated at enrollment and 1, 3, 6, 9, and 12 months after initial diagnosis of T1DM. Preservation of insulin secretion, measured by mixed-meal–stimulated C-peptide secretion, was compared after 6 and 12 months of treatment. Initiation of pump therapy within 1 month of diagnosis resulted in consistently higher mixed-meal tolerance test–stimulated C-peptide values at all time points, although differences were not statistically significant. Improved glycemic control was observed in insulin pump-treated subjects (more time spent with normoglycemia, better mean HbA1c), and pump-treated subjects reported comparatively greater satisfaction with route of treatment administration.

In pregnancy, although data specific to randomized control studies comparing MDI to CSII are not available, for those women who do use insulin in pregnancy, insulin pumps may facilitate normalized glucose with fewer excursions, fewer hypoglycemic episodes, and better adherence to aggressive insulin regimens than with MDI [25]. It is unclear whether pumps provide tighter glucose control or simply ease the use of insulin by reducing the need to carry syringes and insulin, as well as serving as a reminder to monitor and manage blood glucose. Dosing in 1/20ths of units over long and short intervals with the ability to interrupt or modify post hoc allows precision as well as flexibility. Insulin pumps allow for lower insulin doses and may facilitate the use of U-500 regular insulin in the very insulin-resistant woman, although the considerable risk of hypoglycemia with this insulin should be kept in mind. However, there is little evidence to support that insulin pumps improve glycemic control during pregnancy compared to well-managed MDI insulin use [26]. The prevalence of large for gestational age babies and fetal morbidity were similar between groups randomized to insulin pump versus MDI insulin use [27].
Another systematic review by Jeitler [28] included studies of those with both T1 and T2DM. Although only two papers on CSII in T2DM were included, in the patients with T2DM, CSII and MDI treatment showed no significant difference for HbA1c, in contrast to those with T1DM. Severe hypoglycemia was rare, but a trend toward lower incidence was noted in those on CSII. Mild hypoglycemia was not felt to be statistically different in one study, while only a trend toward lower incidence was shown with CSII in the other study. Hyperglycemia events were reported only in one study, with CSII patients having fewer than the MDI patients, but without statistical analysis as to significance being provided, the outcome is nebulous.

Most recently, yet another systematic review and meta-analysis [29] was designed to compare the effects of CSII with MDI. Using randomized controlled trials (RCTs) among children, adolescents, or adults with either T1 or T2DM, or pregnant women with pre-existing diabetes no difference was observed in the effect of CSII and MDI on HbA1c (moderate strength of evidence) or severe hypoglycemia (low strength of evidence) for adults with T2DM. There was insufficient evidence about quality of life for adults with T2DM.

Although randomized control trials showing comparison effects between MDI and CSII are not available, there is some consensus that supports the consideration of CSII over MDI for diabetic patients with certain needs. Specifically, diabetics with changing work schedules or shifts, athletes that benefit from the ability to suspend insulin delivery, those with significant dawn phenomenon or early morning hyperglycemia, patients with gastroparesis as well as those with variable daytime insulin needs. In short, many recommendations for insulin pump therapy are based on best practice approaches.

**Insulin pump use**

While patients are best served by the expertise of an endocrinologist working in collaboration with a certified diabetes educator pump trainer in considering insulin pump therapy, many caveats exist which must be explored in order to glean the maximum benefit of a pump over MDI. One factor to assess is patient motivation as only those with a considerable amount of dedication with benefit from insulin pump therapy. A supportive team is also needed to help with education, initiation, and maintenance of support on insulin pump therapy. Members of the team are relied upon to provide the patient with critical tools (carbohydrate counting, correction insulin algorithms, etc.) all along the process to make CSII successful [30].

Supporting the patient in maintaining their motivation is important as is education on how to best use multiple daily injections. Patients must remain open to learning new strategies for diabetes management in order to be truly successful with insulin pump therapy. Like many aspects of diabetes therapy, and especially insulin pump therapy, there are no data to predict “success” (which can be defined differently by different experts); nevertheless, the need exists for development of some basic steps/skills.

While self-motivation is fundamental to success, without frequent self-monitoring of blood glucose (SMBG) or continuous glucose monitoring (CGM) patients lack necessary information for making adjustments in insulin dosing, food intake, or exercise to decrease glycemic variability. When controlling for frequency of SMBG, there is no advantage for CSII compared to MDI, particularly when there are fewer than four daily blood tests. As the frequency of SMBG increases, there appears an advantage to HbA1c with CSII [31]. While this is simply an observational study and firm conclusions cannot be made, making small insulin changes (often referred to as “fine-tuning” or “micromanaging”) with a pump could potentially have a more favorable impact on HbA1c with more frequent SMBG.

Most insurance companies are now requesting documentation of at least four daily blood glucose tests to better ensure a good outcome with CSII. In our clinic, we strongly encourage a minimum of four SMBG tests each day before consideration of CSII. The use of routine glucose meter downloading ensures ease of documenting frequency of SMBG, which is either sent to the payer electronically, or printed and faxed.

Discussing realistic expectations and goals of CSII is part of building the foundation for insulin pump therapy. Many patients have an unrealistic expectation that their insulin pump will free them from the demands of self-management and ongoing personal attention to their treatment. We often discuss how the opposite is true and that they will be required to spend even more time and attention on diabetes management. From the Joslin group in Boston, MA, it has been noted that those patients who described the pump as a tool, not a cure, to meet glycemic goals, reported a more active approach to their diabetes and had better glycemic control [32].

Knowing what to do with the glucose data from SMBG or CGM is another key factor [30]. Simply measuring glucose levels each day on a standalone basis is not the same as using the data for action. Part of the controversy as to the benefit of SMBG has been that merely checking fingerstick gluoses is not helpful unless it is used to determine an action [33]. Whether using MDI or CSII, patients are best served by knowing what to do with the SMBG information they generate. Ideally all patients using MDI, but particularly those initiating CSII, should understand the basics of correcting pre-meal or between-meal hyperglycemia (and for that matter, hypoglycemia). This should include routine use of correction dose insulin so that patients understand the concepts of “insulin sensitivity factor” (or correction dose), “insulin on board,” and “insulin stacking.” Similarly, understanding the risks of overtreating mild hypoglycemia and how to avoid doing so should be a routine aspect of diabetes education. Diabetes educators are the best resource for teaching these basic concepts to either MDI or CSII users; they are an integral part of the treatment team and their role in insulin pump therapy cannot be emphasized enough [30]. While a physician in solo practice certainly can participate
in all aspects of this type of diabetes education, this is not a realistic option for the majority due to the tremendous time commitment.

Additionally, an understanding of matching insulin to food intake is critical to the success of CSII use. On occasion, a patient may do well without understanding the basics of carbohydrate counting, but this is more the exception. Those individuals who are consistent with their food selections on a daily basis, particularly with frequent SMBG, can often maintain reasonable glucose control by frequent "touch-ups" using small amounts of correction dose insulin or food to correct hypoglycemia. Many patients initiating the process toward CSII have not had any diabetes education, particularly nutrition education, for many years. Consequently, their benefit from meeting with a nutritionist prior to starting insulin pump therapy can be significant. With carbohydrate content now listed on food labels and available in many restaurants, the process of carbohydrate counting and determining insulin doses is far easier than it was in the past. Furthermore, there are many sites on the Internet and even "apps" for smart phones and tablets that can assist patients with understanding carbohydrate content of nonlabeled foods.

Some amount of trial and error in prandial insulin dosing is to be expected and yet there are several general points that many patients, as well as clinicians, do not generally appreciate. For example, if a meal contains a significant amount of fiber (>5 g), one half of the fiber portion of the total carbohydrate value should be subtracted from the carbohydrate dose [34]. So for a 40 g carbohydrate meal with 10 g of fiber, the insulin dose should calculated based on 40 g minus (10 g/2) = 35 g of carbohydrate. In general, exercising within 3 hours of a mealtime bolus or injection means the mealtime carbohydrate ratio should be raised (or a fraction of the recommended dose should be administered) to provide a lower dose of insulin. Finally, the timing of the insulin as it relates to when the meal is consumed (the "lag time") is an underemphasized and under-appreciated aspect of mealtime insulin dosing that should also be reviewed prior to insulin pump therapy [6,35]. Furthermore, confusion exists among both clinicians and patients about the difference between pharmacokinetics and pharmacodynamics with insulin, especially prandial insulin. Those using continuous glucose sensors quickly appreciate the slowness of action of "fast" insulin, as glucose trends are followed in response to a bolus insulin dose.

Even though there may not be agreement on what constitutes success, trying to manage and educate pump patients without a high level of support often leads to failure. The multidisciplinary "diabetes team," has been shown to significantly improve the successful initiation of CSII therapy [7] but, unfortunately, many patients do not have that option. Diabetes care centers in the United States usually have several components of the traditional diabetes team: nurse educators, nutritionists, and occasionally clinical pharmacists, behavioral scientists (psychologists, psychiatrists), advanced nurse practitioners, social workers, and rarely exercise physiologists. These individuals are often certified diabetes educators (CDEs) and are the nucleus that coordinates the glycemic management in all patients, but particularly those using CSII. Many physicians in smaller, non-diabetes centers have some element of the diabetes team in their office or nearby. It is not realistic for patients using CSII to do as well as possible without some assistance in both the initial education and longer term teaching-management. Many communities have CDEs who are knowledgeable about pump therapy, and they are close by, often within a hospital setting for consultation. Access is typically facilitated through a physician referral to ensure reimbursement. Pump companies can also often provide the initial pump training. Between this one-on-one contact and the company's website, many of the initial training issues can be adequately addressed. However, patient follow-through after the initial training is often inconsistent and so pump management suffers. Pumps offer many advanced features which do not get covered in the initial training so by failing to maintain contact, patients do not reap the full reward of insulin pump therapy such as bolus calculators and temporary basal rates. Finally, much of the initial education can be forgotten if it is not being actively reinforced by the patient's primary diabetes provider. In our experience, the patient's old habits will return and the pump becomes a more frustrating than helpful tool with diabetes management.

Nontraditional options do exist for patients where the physician (or nurse practitioner) does not have access to these other components of the diabetes team. Pump manufacturing companies are endeavoring to fill the gap by providing advanced pump training classes and support groups. Depending on the pump chosen, one may find in-depth training via a sophisticated and interactive website. There are also outstanding reference books and apps for smartphones, with websites for all insulin pump companies that allow for some hands-on experience in managing insulin pumps "virtually." Some even provide a step-by-step introduction of pump use.

Thankfully, the traditional clinic setting is not the only place to gain the much-needed education and training.

Real-time continuous glucose monitoring (RT-CGM) linked to insulin pump

As noted earlier, advances in combining continuous glucose sensing with pump therapy is moving technology toward the ultimate goal of an artificial pancreas. Data have been accumulating on the integrated modalities of RT-CGM and pumps specific to glycemic control, quality of life, and, hopefully in the near future, on diabetes complications prevention. Early studies, such as those of Peyrot et al. [36] focused on both patient acceptance of an integrated system, as well as effects on glycemic control. In this study, 28 patients with T1DM with HbA1c from 7.5–11.1% (average 8.6%) were randomized to MDI and SMBG (control group) or CSII and RT-CGM
(study group) and results reviewed after 4 months. Parameters examined included weight, HbA1c, severe hypoglycemia, and episodes of ketoacidosis. Additionally, a number of satisfaction surveys were completed targeting the integrated system users and user acceptance. Outcome data showed statistically significant decreases in HbA1c from baseline of 1.7% in the study group (8.87 ± 0.89% to 7.16 ± 0.75%; p ≤ 0.001) versus 1.0% in the controls (8.32 ± 1.05% to 7.30 ± 0.92%; p ≤ 0.002); however, the difference in change between the two groups (0.7%) was not statistically significant (p = 0.071). There were small nonsignificant increases in weight from baseline of 0.7 kg in the study group and 2.0 kg in the controls; the difference in change between groups (1.3 kg) was not significant (p = 0.309). There were three severe hypoglycemia events, one incidence of diabetic ketoacidosis, all in the controls. Of more interest, the integrated RT-CGM/insulin pump system was associated with substantial patient-reported benefits over SMBG/MDI. Treatment satisfaction measures, particularly burden of blood glucose monitoring and convenience, as well as measures of health-related quality of life, including social burden and diabetes-related worries were significantly more positive for the integrated system uses. However, there was no difference in measures regarding interference in daily activities and psychological well-being, despite which, the integrated system users were less likely to want to switch away from RT-CGM/CSII and were more likely to recommend it to others.

A larger study by the same group [37] assessed satisfaction between CSII and SMBG versus CSII and RT-CGM. Just over 300 patients with median age 43 years, majority (94%) with T1DM, majority with diabetes duration over 15 years (61%), and median use of insulin use 15 years were queried through an internet survey. The RT-CGM/CSII group gave significantly better ratings than the SMBG + CSII group for their glucose monitoring system’s glucose control efficacy, overall satisfaction, desire to switch, and willingness to recommend, and significantly worse ratings for interference with daily activities. The RT-CGM/CSII group gave significantly better ratings than the SMBG + CSII group for their insulin delivery system’s convenience and glucose control efficacy, overall satisfaction, desire to switch, and willingness to recommend. These results support patient satisfaction with RT-CGM/CSII.

The JDRF trial [38] provided outcome data regarding combined RT-CGM and CSII use in both children and adults. Three hundred and twenty-two study participants from multiple centers already receiving intensive therapy for T1DM were randomized to a group with RT-CGM or to a control group performing SMBG. HbA1c at baseline ranged from 7 – 10%, ages ranged from 8 years to over 25 years. The primary outcome was change in the HbA1c at 26 weeks. Although this study included participants who were on MDI, the majority were reported as using CSII, with HbA1c of 8.0% or less, and measuring glucose levels more than five times daily. All three available RT-CGM systems were used in this study.

Results showed a clear impact in decreased HbA1c with use of RT-CGM, irrespective of whether MDI or CSII was used. Results were stratified by age, with the over 25 year olds achieving most significant HbA1c lowering, mean change, −0.53%; 95% CI −0.71 to −0.35; p < 0.001). In primary analyses, neither those 15–24 years of age nor those 8–14 years of age had statistical changes in HbA1c. However, the 8–14 year olds started on RT-CGM were more likely to have had a baseline HbA1c of less than 7.0%, and more had a decrease of 10% or more from baseline.

There was also a clear relationship between the frequency of use of RT-CGM and improved HbA1c; furthermore, the group that wore the RT-CGM the least was the age group that did not see an HbA1c impact. Severe hypoglycemic events were infrequent in both study groups, with no significant differences between the groups on the basis of age. This is significant, in that with the HbA1c improvement in the over 25 year olds, there was no associated increase in hypoglycemia—in direct contrast to the DCCT [39] which clearly showed a significant increase in hypoglycemia with lower HbA1c levels.

STAR 3 [40] focused entirely on an integrated RT-GM and CSII protocol in study participants. Over 400 participants from multiple centers were randomly assigned to receive either sensor-augmented pump therapy or a regimen of multiple daily injections, stratified according to age group: adults (19 – 70 years of age) or children (7 – 18 years of age), and followed for 1 year. HbA1c was measured, as well as rates of severe hypoglycemia and ketoacidosis. Baseline mean HbA1c in both intervention as well as control groups of 8.3% decreased to 7.5% in the pump-therapy group (absolute reduction, 0.8 ± 0.8 percentage points), as compared with 8.1% in the injection-therapy group (absolute reduction, 0.2 ± 0.9 percentage points), for a between-group difference in the pump-therapy group of −0.6 percentage points. Both age groups benefitted, in the adult group, absolute reduction in the mean HbA1c was 1.0 ± 0.7 percentage points in the pump-therapy group, and 0.4 ± 0.8 percentage points in the injection-therapy group, for a between-group difference in the pump-therapy group of −0.6 percentage points. Among children, an absolute reduction in HbA1c of 0.4 ± 0.9 percentage points in the pump-therapy group but an increase of 0.2 ± 1.0 percentage points in the injection-therapy group, showed a between-group difference favoring the pump-therapy group of −0.5 percentage points.

There was no difference in rates of severe hypoglycemia or ketoacidosis between groups, or by age. Additionally, frequency of sensor use of 41 – 60% was associated with a reduction of 0.64 percentage points in HbA1c levels, and increasing sensor use to more than 80% doubled the effect. The difference between the JDRF study and STAR 3 trial results specific to use of sensor and associated HbA1c improvement was suggested as attributed to therapy management software used in the latter.

Most recently the combination of a continuous glucose sensor and pump delivery system that has the added capacity for insulin delivery suspension at a preset glucose threshold has advanced
Adding RT-CGM to pump therapy

As with the initiation of insulin pump therapy, RT-CGM initiation requires a motivated patient, discussion of expectations, a team that supports the educational process for successfully using RT-CGM, as well the initiation and management of data stages following the start of RT-CGM. Presently, our clinic prescribes an initial presensor class, then one-on-one initiation, followed by prompt and routinely scheduled appointments to cover ongoing management. As indicated, patients also receive telephone support to provide timely education between appointments. Evert and colleagues from our clinical center [42] have outlined a program for RT-CGM starting with an instructional presentation in class setting, after RT-CGM use is recommended by a clinician to a patient. There is limited data regarding which patients may be best suited for using real-time CGM as a self-management tool. Ritholz and others [32,43] attempted to identify patient groups more likely to benefit from insulin pump initiation. Reportedly, patients with higher A1c values expected the device to improve glycemic control without their active involvement, whereas patients with lower A1c values understood that they were the agents of change and conceptualized the insulin pump as a tool that assisted their self-care. Predictably, active participation in self-care and realistic expectations of the device were associated with better glycemic control. Similarly, it would be expected that the same would hold true for any diabetes technology, and specifically a continuous glucose sensor system. By adding the RT-CGM, you must also add education beyond grasping elements of diabetes knowledge and demonstrating the required skills for pump initiation. Additional understanding is necessary about how the RT-CGM operates and the fact that it relays glucose trends and not absolute values. This point is critical, as patients need to be informed about the likely need to increase SMBG in order to validate true glucose levels when glucose trends show steep rises or declines due to the delay in sensor readings from interstitial fluid reflecting actual glucose levels.

While the structured approach to RT-CGM offered in our clinic has yet to be validated, our center’s experience reflects less patient frustration with insertion technique and increased confidence in how to respond to RT-CGM obtained data. We do recommend a two-step process for patients working toward the use of the integrated sensor and pump system: (1) obtain a pump and develop pump-use skills; (2) add the RT-CGM component. Making use of this approach results in very few patients being overwhelmed whereas starting with both components has led to considerable frustration and in some cases outright refusal to persist with either modality. There are always exceptions, when this process works as well with the sensor first, followed by the pump. Studies as to best approach are clearly needed.

Following the presensor class, at which the various sensors are presented and discussed as to operational details, patients are given contact information should they decide to place an order. Empowered with the option to follow through, the patient is either accepting or declining the addition of a sensor to their pump. As patients obtain the RT-CGM system, they are called with instructions on how to prepare for their scheduled initiation. One requirement is to charge the system transmitter or receiver the night before and review the manufacturer’s instruction materials. Educators again play an important role during initiation as they explain how RT-CGM differs from capillary SMBG and reinforce the need for continued SMBG for the ongoing calibration of the system.

In the beginning, patients should be encouraged to note the information from the RT-CGM receiver for 3 to 7 days without taking action in order to reduce the potential for overcorrection in response to alarms. It is important to inform patients that it is very common to become overwhelmed and frustrated by the amount of data and information they will be receiving in “real time” [44]. Another norm with real-time CGM is that patients will be out of target range more than 50% of the time [44]. Informing patients on what to expect will help minimize their frustration and worry while maximizing the effectiveness of the modality. Following-up soon after RT-CGM initiation is sound practice to resolve technical problems such as insertion technique, offer reminders regarding calibration with SMBG, reevaluate settings specific to high and low glucose alarms, and perhaps most importantly, to review. Each patient has a unique learning curve and will need guidance on how to effectively use the data they are seeing to adjust insulin dosing, timing of dosing, meal content or meal amount, or physical activity.

Future trends

Even with all of the technological advances in insulin pumps and their integration with other modalities, barriers to its adoption still exist for patients, besides cost. Improving many features of the external pump could reduce not only treatment constraints, but also improve quality of life. Drawbacks include: initial technical education on how to use the pump and insert the catheter takes time; the negative impression some patients have of being attached to an external device, and common equipment problems, such as catheter occlusion and bent cannula; time-consuming recommendations to disconnect the pump before taking a shower or engaging in water or other sports activities [45]. Upgrades to infusion sets are indicated to achieve the goal of reducing insulin occlusions and unexplained hyperglycemia.

Patch pumps have appealed to some, as they are smaller and lighter than pumps with cannulas and are free of tubing making them more discrete. Patients require less instruction time on
their use and the device is usually simpler in nature than the typical pump thereby reducing technical issues. A few of these pumps are already available for use; others will be in the near future [46]. Some have described the patch pump as an insulin pen that is worn on the body. The basic design comprises an insulin reservoir, delivery system and cannula, all of which are integrated into a small, wearable, disposable or semidisposable device. For some models, insulin delivery can be directed though a separate handheld personal data manager (PDM), for others, insulin is delivered through a preset basal rate, but allowing an on-demand bolus delivery, changed daily. Other models in development allow for the handheld data manager to not only deliver insulin, but contain a glucose meter and food library. Yet another patch pump will increase the reservoir capacity so that refills occur every 6 days with the pump lasting 3 months. Many of these latter pumps are being targeted to the market for T2DM since that population is more suited to the limited basal rate adjustment capacity of the patch pump [45].

Next on the horizon is a closed loop completely integrated insulin delivery and glucose-sensing system. Hypoglycemia sensing remains the major concern for safety in a closed loop system. Alerts for this condition are integrated into RT-CGM systems; however, the DirectNet Study Group [47] showed that 71% of patients—specifically, children and adolescents—did not react to the hypoglycemia alerts that occurred during sleep. This is particularly concerning as most episodes of severe hypoglycemia happen at night [48]. In response, the low-glucose suspend (LGS) system has been developed and is in use. The Veo™ model (marketed in the United States as the Medtronic 530 pump system with Enlight sensor), system uses data transmitted by RT-CGM sensor to automatically suspend the delivery of insulin in cases of hypoglycemia. This “automatic stop” function is activated when the sensor detects interstitial glucose levels below a predetermined level. Clinical experience with the Paradigm Veo™ pump is still limited, and the efficacy of the system in reducing hypoglycemia has only been evaluated in a few recent short-term studies [41,49]. In the Choudhary study [41], hypoglycemia events in those with T1DM, were examined during two consecutive periods. During the first 2-week period, the automatic stop was not activated (LGS was set on OFF) During the consecutive 3 weeks, the LGS was enabled (set on ON). The results showed that patients with a high risk of hypoglycemia significantly reduced their nighttime hypoglycemia duration with no hyperglycemic rebound or ketosis while using the LGS function. Similar results were reported in the Danne study [49] in diabetic children and adolescent subjects. Finally, Garg et al. [50] showed a 19% reduction of hypoglycemic exposure with exercise-induced hypoglycemia.

Reducing hypoglycemia is a central goal in the overall balance of pros and cons to any diabetes therapy. Intraperitoneal delivery of insulin has the advantage of approximating more physiologic insulin action kinetics. Intraperitoneal insulin is absorbed mainly in the portal vein, allowing for lower peripheral insulin levels and even improved glucagon response to hypoglycemia [51]. The use of the intraperitoneal route for T1DM treatment was made possible by the development of programmable implantable pumps that deliver insulin through an intraperitoneal catheter. As an example, the implanted MIP 2007 model (Medtronic-MiniMed, Northridge, CA, USA), was first in use in 2000, having a 7- to 10-year battery life. Insulin delivery options are similar to those of the most up-to-date external pumps, and are programmable through a personal pump communicator (PPC). A patient using a Medtronic implantable insulin pump reported his positive experiences in 2008, offering 8 years of observations of use. With delivery of insulin directly into the peritoneal cavity, he reported feeling better, having more flexibility in eating, and fewer insulin reactions. However, at the conclusion of the article, he expresses disappointment that the manufacturer of the pump, Medtronic, no longer plans to seek Food and Drug Administration approval for this therapy [52].

The metabolic benefits, from observational studies on implantable pumps, show a decrease in HbA1c, a decrease in the frequency of severe hypoglycemia, and even less glycemic variability [45]. These benefits persist over time, even in type 1 diabetics who have not even closely reached HbA1c target of 7% and/or have large blood glucose fluctuations, severe recurrent hypoglycemia, despite tight coaching and intensified education with previous subcutaneous insulin treatment [53]. A Dutch study [54] has shown that, with an implanted insulin pump, not only was HbA1c significantly improved in those who were previously poorly controlled, but instability-related diabetic hospitalizations were also significantly reduced.

However, despite reported improvement in quality of life [46], use of implantable pumps remains limited, essentially prescribed only in Europe. The limitations of this treatment mode are the result of its technically specialized medical requirements, significant cost with lack of reimbursement, as well limited manufacturing. With regard to implanted pumps coupled with glucose sensors, the pharmacokinetic properties of intraperitoneally administered insulin is of particular interest for use in a closed-loop system. Pilot studies have shown encouraging results with implanted pumps coupled with intravenous [55] and more practical subcutaneous glucose sensors [56].

**Conclusion**

Progress has been both steady and encouraging since Kadish’s first pump prototype was unveiled in the 1960s. Safer, smaller, smarter pumps now coupled to continuous glucose sensing have given patients a more comprehensive diabetes management tool that takes advantage of computerized programs for tracking insulin doses and integrating glucose and carbohydrate data to guide patients in improving their care. The resident intelligence of these devices has taken great strides to where they provide suggestions for bolus insulin dosing and even turn themselves...
<table>
<thead>
<tr>
<th>Manufacturer</th>
<th>Animas</th>
<th>Medtronic MiniMed</th>
<th>OmniPod®</th>
<th>Roche</th>
<th>Tandem® Diabetes Care</th>
<th>Valeritas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>One Touch™ Ping™</td>
<td>MiniMed® 530G with Enlite®</td>
<td>Diabetes Insulin Pump</td>
<td>ACCU-CHEK® Spirit Insulin Pump</td>
<td>t:slim® Insulin Pump</td>
<td>V-Go™ Insulin Delivery System (technically not a “pump”)</td>
</tr>
<tr>
<td>Dimensions</td>
<td>3.25 x 2 x 0.86</td>
<td>3.6 x 2 x 0.8 inches</td>
<td>1.6 x 2.4 x 0.7 inches</td>
<td>3.2 x 2.2 x 0.8 inches</td>
<td>3.13 x 2 x 0.6 inches</td>
<td>2.4 x 1.3 x 0.5 inches</td>
</tr>
<tr>
<td>Weight</td>
<td>&lt; 4 oz. full</td>
<td>1.2 oz. full</td>
<td>1 oz. full</td>
<td>3.95 oz. full</td>
<td>3.75 oz. full</td>
<td>0.7 to 1.8 oz.</td>
</tr>
<tr>
<td>Basal increments</td>
<td>0.025–25 U h⁻¹</td>
<td>0.025–35 U h⁻¹</td>
<td>0.05–30 U h⁻¹</td>
<td>0.1–25.0 U h⁻¹</td>
<td>0.1–15.0 U h⁻¹</td>
<td>2 units increments up to 36 total in 24 h</td>
</tr>
<tr>
<td>Reservoir capacity</td>
<td>200 units</td>
<td>300 units</td>
<td>200 units</td>
<td>315 units</td>
<td>300 units</td>
<td>3 sizes: 56, 66, and 76 unit capacity (basal plus 36 unit bolus)</td>
</tr>
<tr>
<td>Basal patterns</td>
<td>12</td>
<td>7 programs with up to 24 segments each</td>
<td>5 basal rate profiles</td>
<td>4 basal profile programs</td>
<td>3 sizes or basal rates: 20(0.83 U h⁻¹), 30(1.25 U h⁻¹), 40(1.67 U h⁻¹)</td>
<td></td>
</tr>
<tr>
<td>Insulin on board calculation</td>
<td>Yes</td>
<td>Linear degradation. Tracks correction boluses</td>
<td>5 basal rate profiles</td>
<td>Yes</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Special features</td>
<td>Meter remote</td>
<td>Built-in CGM, automatic pump suspend when threshold glucose reached, unless user cancels</td>
<td>7 programs with up to 24 segments each</td>
<td>Tubeless and wireless pump</td>
<td>Reversible screen Backup pump</td>
<td>Uses one fast-acting insulin (Humalog® or NovoLog®)</td>
</tr>
<tr>
<td></td>
<td>Color screen is self-illuminating</td>
<td>Always tracks active insulin in bolus calculations</td>
<td>Programmable in 30-min increments</td>
<td>Integrated BG meter</td>
<td>The pump can wirelessly download to a handheld device or computer for analysis</td>
<td>No tubing or cannula</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Contour Next Link Meter that transmits BG result to pump and downloads pump to Carelink</td>
<td>Temporary basal rate in 10% increments from 0% to 200%, and 15-min increments from 15 min to 24 h</td>
<td>Integrated food library</td>
<td>Automatic cannula insertion</td>
<td>No need to plan your meals on an insulin schedule for mealtime bolus dosing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lowest available insulin to carb ratio of 1:1</td>
<td>Range 0% to 250%</td>
<td>Automatic cannula insertion</td>
<td>Continuous Insulin Delivery; no need to ever disconnect from your pump</td>
<td>The V-Go buttons can be pressed through your clothes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Capture events feature</td>
<td>Missed meal bolus reminder</td>
<td>BG reminder</td>
<td>Reversible screen Backup pump</td>
<td>Flexibility to choose a new V-Go application site every 24 h to work with your clothing</td>
</tr>
<tr>
<td><strong>Pump programming</strong></td>
<td>Carelink Professional and Carelink Personal Software</td>
<td>Personal Diabetes Manager (PDM) CoPilot® Health Management System</td>
<td>Pocket Compass software t:connect™ Therapy Management Software</td>
<td>None. Preset basal rates</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>-----------------------------------------------------</td>
<td>-----------------------------------------------------------------</td>
<td>------------------------------------------------</td>
<td>-----------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Bolus &quot;types&quot;</strong></td>
<td>4 options</td>
<td>Easy Max Normal Square wave Dual wave</td>
<td>Standard Extended Combination Plus 7 personalized bolus presets</td>
<td>Quick Scroll Extended MultiWave Pre-programmable bolus increments of 0.1, 0.2, 0.5, 1.0, 2.0 U</td>
<td>Food bolus Extended bolus</td>
<td></td>
</tr>
<tr>
<td><strong>Waterproof</strong></td>
<td>Yes</td>
<td>No—water-resistant</td>
<td>Yes, 25 feet up to 60 min</td>
<td>Yes</td>
<td>Yes—watertight</td>
<td></td>
</tr>
<tr>
<td><strong>Warranty</strong></td>
<td>4 years</td>
<td>4 years (pump) 6 months (CGM)</td>
<td>4 years</td>
<td>4 years</td>
<td>4 years</td>
<td></td>
</tr>
<tr>
<td><strong>Battery</strong></td>
<td>1 AA alkaline battery</td>
<td>AAA alkaline battery</td>
<td>2 alkaline AAA batteries</td>
<td>1 AA alkaline battery</td>
<td>Re-chargeable</td>
<td></td>
</tr>
<tr>
<td><strong>Colors</strong></td>
<td>Blue, black, silver, pink, green Clear, purple, pink, blue, smoke</td>
<td>PDM skins: red, yellow, green, pink, purple, black, clear</td>
<td>Skins in blue, electric blue, black, pink, yellow, and white. Stickers in 16 patterns</td>
<td>Black with silver trim</td>
<td>Clear opaque</td>
<td></td>
</tr>
</tbody>
</table>
off completely when a low glucose level is perceived. Still, no system is perfect and pumps still occlude, sensors can “break,” and the transmission of data between devices can be disrupted thereby causing problems for users. As technology advances the options for diabetes management, providers must also stay in step to ensure that patients feel comfortable embracing those options. In total, the future holds much promise toward continued improvements in a system that will eventually take over the management of glucose and insulin for everyone with diabetes mellitus. A table of current insulin pump models can be found at the end of this article (Table 49.1).

Acknowledgment

A special acknowledgement to Elisa Washington for her invaluable assistance and wonderful help in the preparation of this manuscript.

References

28 Jeitler K, Horvath K, Berghold A, et al.: Continuous subcutaneous insulin infusion versus multiple daily insulin injections in patients


Innovative therapies in diabetes: colesevelam and bromocriptine

Shamsa Ali and Vivian Fonseca
Section of Endocrinology, Tulane University Health Sciences Center, New Orleans, LA, USA

Key points

• There is an ongoing need for development of novel therapies for the treatment of type 2 diabetes mellitus. Two oral agents (colesevelam and bromocriptine) recently became available for the treatment of T2DM.
• Colesevelam is an antihyperlipidemic agent approved by the US FDA for the treatment of T2DM. The ability of colesevelam to lower LDL-cholesterol and hyperglycemia makes it a useful single agent to address both hyperlipidemia and hyperglycemia in patients with T2DM.
• Colesevelam has a good safety profile as it is not absorbed systemically; however, it should be used with caution in patients with hypertriglyceridemia.
• Bromocriptine-QR, a dopamine receptor (D2) agonist is a first-in-class agent that is FDA approved for the treatment of T2DM. Bromocriptine has a unique mechanism of action that is unlike any other medication currently available for the treatment of T2DM.
• Bromocriptine-QR works centrally by resetting the hypothalamic circadian rhythm and increasing central dopamine activity.
• Bromocriptine-QR is a safe and efficacious treatment for T2DM.

Colesevelam

Colesevelam HCL is a nonsystemic bile acid sequestrant (BAS) approved by the FDA for the treatment of dyslipidemia and T2DM. Treatment with colesevelam is associated with significant reductions not only in LDL-cholesterol (LDL-C) but also in fasting plasma glucose and hemoglobin A1c. Colesevelam is especially beneficial in the management of patients with T2DM and elevated LDL-C. Caution should be observed in patients with hypertriglyceridemia, as treatment with colesevelam is associated with a significant increase in triglyceride levels. Trivial reductions in total cholesterol and high-density lipoprotein (HDL) have been reported.

Mechanism of action

Bile acids are synthesized in the liver, transported to the duodenum, and are essential for digestion as well as absorption of fats in the food. Approximately 90% of the bile acids are reabsorbed in the terminal ileum and enter into enterohepatic circulation. Colesevelam interferes with enterohepatic circulation by binding to the bile acids and forming complexes to be excreted in feces. This causes deficiency in bile acids, which in turn stimulates an increase in the synthesis of bile acids in the liver utilizing cholesterol and thus decreasing the intrahepatic cholesterol concentration. This leads to upregulation of LDL receptors on the liver surface to help preserve hepatic cholesterol balance resulting in increased clearance of LDL from blood leading to decreased serum concentration of LDL and explains the cholesterol-lowering effect of colesevelam as well as other BAS [1].

Mechanism of glycemic improvement

The mechanism by which bile acid sequestrants affect glycemia is unclear. One suggested mechanism of action is through their effect on various nuclear receptors. Farnesoid X receptor (FXR), which is present in both liver and intestines, liver X receptor (LXR), which is present in liver, and a G protein-coupled receptor (TGR-5) present in the intestines are implicated in the regulation of glucose metabolism and also have a crucial role in bile acid and triglyceride homeostasis [2–4].

Bile acids exert a negative feedback effect by activating FXR to prevent build-up of high levels of bile acids in the liver [3]. FXR possibly regulates glucose metabolism via a pathway mediated by fibroblast growth factor 15/19 (FGF15 (mice)/19(humans)). Increased metabolic rate and decreased adiposity have been reported in transgenic mice expressing human FGF19 [5].
Treatment with a bile acid sequestrant in patients with T2DM and hypercholesterolemia has shown a significant increase in GLP-1 levels explaining at least in part the mechanism of glycemic improvement with the use of BAS [6].

Once FXR is activated, it upregulates expression of a small heterodimer protein (SHP) which is an inhibitor of LXR. LXR-induced bile acid synthesis is inactivated by the SHP [3].

**Pharmacokinetic and pharmacodynamics**

Pharmacokinetic data with all BAS is limited mainly because they are not systemically absorbed [7]. Colesevelam undergoes nonsignificant systemic absorption and also has minimal tissue concentration, volume of distribution, plasma protein binding, and systemic metabolism. It has minimal renal clearance due to negligible systemic absorption and is mostly excreted through feces [8,9]. Colesevelam has a higher potency and its binding affinity is four to six times greater than other BAS [7].

**Efficacy of colesevelam**

**Low-density lipoprotein**

Several classes of cholesterol-lowering drugs, namely statins, nicotinic acid, fibrates, cholesterol absorption inhibitors, and BAS are used for treatment of dyslipidemia. It is well known that BAS lower LDL-C [10,11] (Table 50.1). A meta-analysis of pooled data from eight studies reported a statistically significant reduction in LDL associated with colesevelam [12].

Colesevelam can be used as a monotherapy as well as in combination with other lipid-lowering agents. In a double-blind placebo-controlled trial, colesevelam as a monotherapy lowered LDL-C by up to 18% [13] whereas when combined with statin it has been shown to reduce LDL-C by as much as 48% [14].

**Hemoglobin A1c**

Treatment with colesevelam is associated with a statistically significant reduction in hemoglobin A1c [12] (Table 50.2). In a short-term double-blind crossover trial, cholestyramine (bile acid sequestrant) was reported not only to improve dyslipidemia but also glucose metabolism [15]. This observation led to the Glucose-Lowering Effect of WelChol Study (GLOWS), which evaluated the glucose-lowering effects of colesevelam HCl in patients with inadequately controlled T2DM [16]. Colesevelam added to an existing metformin and/or sulfonylurea regimen was associated with significant reductions in A1c, fructosamine, and postprandial blood glucose level. Reduction in A1c of up to 1.0% was reported in patients with higher baseline A1c (at least 8%) [16].

In a 26-week, randomized, double-blind, placebo-controlled study colesevelam was added to the regimen of patients with inadequately controlled T2DM who were on metformin alone or metformin combined with other oral agents. The study reported significant lowering of the mean HbA1c level compared with placebo (−0.54%; *p* < 0.001) [17].

### Table 50.1 LDL mg dL⁻¹

<table>
<thead>
<tr>
<th>Study</th>
<th>Baseline</th>
<th>Colesevelam (C)</th>
<th>Placebo (P)</th>
<th>Treatment difference (95% CI)</th>
<th><em>p</em> value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bays et al. [17]</td>
<td>Mean baseline SD 105.6 (33.79) C 99.0 (28.96) P</td>
<td>101.7 (32.83)</td>
<td>91.7 (39.12)</td>
<td>101.7 (32.83)</td>
<td>(−21.7% to −10.2%)</td>
</tr>
<tr>
<td>Devaraj et al. [50]</td>
<td>150 ± 33C 158 ± 33P</td>
<td>149 ± 27</td>
<td>136 ± 37</td>
<td>149 ± 27</td>
<td><em>p</em> &lt; 0.001</td>
</tr>
<tr>
<td>Davidson et al. [51]</td>
<td>200 ± 39C 194 ± 25P</td>
<td>193 ± 29</td>
<td>181 ± 52</td>
<td>At 6 wks (3 g d⁻¹)</td>
<td>With treatment −18 ± 35</td>
</tr>
<tr>
<td>Garg et al. [52]</td>
<td>ANOVA models adjusted for baseline LDL ↔</td>
<td>At 12 wks −9.6% (95% CI: −16.0 to −3.1)</td>
<td>−1.4% (95% CI: −7.8 to −5.0)</td>
<td>−12.3 (2.35)</td>
<td><em>p</em> = 0.005</td>
</tr>
<tr>
<td>Goldberg et al. [53]</td>
<td>Mean baseline SD 101.5 (27.70) C 101.5 (29.08) P</td>
<td>0.5 (2.40)</td>
<td>At 6 wks</td>
<td>−12.3 (2.35)</td>
<td>[−19.3 to −6.3]</td>
</tr>
<tr>
<td>Insull et al. [54]</td>
<td>130 to 220 mg dL⁻¹ C 4.5 g d⁻¹</td>
<td>Median ↓20%</td>
<td>Mean ↓9 to 18%</td>
<td>C 4.5 g d⁻¹</td>
<td>Change observed at 2 wks lasted till the end of the study at 24 wks</td>
</tr>
<tr>
<td>Zieve et al. [16]</td>
<td>Baseline to 12 wks 182.1 (96.7) C 213.7 (306.2) P</td>
<td>228.6 (110.2)</td>
<td>206.9 (139.9)</td>
<td>228.6 (110.2)</td>
<td>[−19.7 to 35.3]</td>
</tr>
</tbody>
</table>

C, colesevelam; P, placebo; SD, standard deviation
Chapter 50

Table 50.2 HbA1c

<table>
<thead>
<tr>
<th>Study</th>
<th>Baseline HbA1c</th>
<th>Colesevelam Mean (SD)</th>
<th>Placebo Mean (SD)</th>
<th>Treatment difference (95% CI)</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bays et al. [17]</td>
<td>8.1 (0.65) Colesevelam 8.1 (0.61) Placebo</td>
<td>7.7 (1.07)</td>
<td>8.3 (1.09)</td>
<td>(−0.76 to −0.32)</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td>Garg et al. [52]</td>
<td>ANOVA models adjusted for baseline A1c ↔ At 4 wks −0.17 (95% CI: −0.31 to −0.02)</td>
<td>0.03</td>
<td>0.03</td>
<td>[−0.68 to −0.32]</td>
<td>p = 0.001</td>
</tr>
<tr>
<td>Goldberg et al. [53]</td>
<td>Mean baseline SD 8.3 (0.61) Mean change in HbA1c level (percentage) 8.2 (0.62) From baseline to week 16</td>
<td>7.8 (0.84)</td>
<td>8.3 (1.01)</td>
<td>[−0.9 to −0.2]</td>
<td>p = 0.007</td>
</tr>
<tr>
<td>Zieve et al. [16]</td>
<td>Baseline 12 wks 7.9 (0.8) C 8.1 (0.9) P</td>
<td>7.7 (0.6)</td>
<td>8.3 (1.2)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

C, colesevelam; P, placebo; SD, standard deviation

Triglycerides
Treatment with colesevelam is associated with a statistically significant increase in triglycerides [12] (Table 50.3). Hence its use is contraindicated in patients with triglyceride levels >500 mg·dL⁻¹ and a history of triglyceride-induced pancreatitis [18].

High-density lipoprotein and total cholesterol
Colesevelam therapy has been reported to cause a statistically insignificant reduction in HDL and total cholesterol [12].

Safety of colesevelam
The most common side effects are related to the gastrointestinal tract including flatulence, constipation, and dyspepsia. The incidence of constipation was reported to be 8.7–11%. Patients who have preexisting constipation can develop fecal impaction [18].

Hypertriglyceridemia is a known side effect of colesevelam as with other BAS. Increase in serum TG concentrations by a median of 5% has been reported in clinical trials. Lipid parameters should be monitored prior to starting colesevelam [18].

Transient hypoglycemia of mild to moderate severity has been reported when colesevelam HCl in used in combination with metformin, sulfonylureas, or insulin in adults with inadequately controlled T2DM [19].

Colesevelam should be used cautiously in patients with a fat-soluble vitamin deficiency. Animal data suggest the possible development of hemorrhage due to colesevelam-induced

Table 50.3 Triglycerides

<table>
<thead>
<tr>
<th>Study</th>
<th>Baseline TGs</th>
<th>Colesevelam</th>
<th>Placebo</th>
<th>Treatment difference (95% CI)</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bays et al. [17]</td>
<td>172.3 (102.0) C 166.0 (114.3) P</td>
<td>188.0 (127.0)</td>
<td>176.0 (136.5)</td>
<td>(−3.9 to 13.4)</td>
<td>p = 0.22</td>
</tr>
<tr>
<td>Devaraj et al. [50]</td>
<td>127 (92, 200) C 114 (88, 147) P median (25th, 75th percentiles)</td>
<td>159 (108, 203)</td>
<td>126 (78, 151)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Davidson et al. [51]</td>
<td>137 ± 70C 169 ± 78P</td>
<td>At 6 wks (3 g·d⁻¹) 144 ± 66</td>
<td>157 ± 95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Garg et al. [52]</td>
<td>ANOVA models adjusted for baseline A1c ↔ 19.2 (4.1 to 34.3)</td>
<td>7.7 (−7.1 to 22.4)</td>
<td></td>
<td>p = 0.28</td>
<td></td>
</tr>
<tr>
<td>Goldberg et al. [53]</td>
<td>Mean baseline SD 155.0 (108.0) C 167.0 (105.0) P</td>
<td>22.7 (36.7)</td>
<td>0.3 (33.7)</td>
<td>[12.4 to 30.1]</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td>Zieve et al. [16]</td>
<td>Baseline to 12 wks 122.6 (32.7) C 119.6 (26.2) P</td>
<td>107.8 (27.5)</td>
<td>121.8 (28.5)</td>
<td>[−20.2 to −3.3]</td>
<td>p = 0.007</td>
</tr>
</tbody>
</table>

C, colesevelam; P, placebo; SD, standard deviation
vitamin K deficiency. To ensure adequate absorption oral vitamin supplements should be taken at least 4 h prior to taking colesevelam [18].

**Drug interactions**

Drug interactions are not expected with colesevelam as it is not systemically absorbed; however, due to its positive charge it can bind with co-administered drugs in a nonspecific manner reducing their bioavailability.

The structure of colesevelam allows it use for greater bile acid-binding potential in contrast to other BAS, possibly minimizing the potential for drug–drug interactions [20].

Oral contraceptives containing ethinyl estradiol and norethindrone, levothyroxine, and phenytoin are known to interact with colesevelam and should be administered ~4 h before colesevelam [7].

**Conclusion**

Colesevelam causes significant reductions in hemoglobin A1c, plasma fasting glucose, and LDL-C and hence is useful in the treatment of patients with elevated LDL-C and for suboptimal glycemic control in patients with T2DM. Colesevelam can be used by itself or added to an existing regimen for further management of hyperglycemia and hypercholesterolemia. The dual glucose- and lipid-lowering effect of colesevelam makes it unique and helps achieve both glycemic and lipid goals in patients with T2DM.

Attention should be paid to the baseline triglyceride level, as use of colesevelam is associated with an increase in triglyceride level. Colesevelam is nonsystemic and well tolerated.

**Bromocriptine-QR**

Bromocriptine-QR can be used by itself or in addition to all other available treatments for T2DM. It is a treatment modality with a novel mechanism of action; although not fully understood, it works centrally by resetting hypothalamic circadian rhythm. It delivers a distinct and short-lived interval of circulating bromocriptine in the morning, increasing central dopamine activity.

**Mechanism of action**

Bromocriptine-QR mediates its effects on glucose metabolism by not directly acting on specific receptors; instead its effects are mediated by resetting the central nervous system (CNS) sympathetic and dopaminergic tone [21]. Addition of bromocriptine-QR to usual treatment regimens for T2DM has shown improvement in fasting as well as postprandial plasma glucose. It is safe to use and is associated with fewer cardiovascular outcomes. Addition of bromocriptine-QR to the existing treatment plan for T2DM provides a new tactic to address inadequately controlled T2DM.

The idea of using bromocriptine-QR for the treatment of T2DM emerged while studying the metabolism of hibernating animals and migrating birds [21,22]. They alter their metabolism according to the time of the year from insulin-sensitive to insulin-resistant states to survive long periods of food deprivation during winter. During this transition there is an increase in lipolytic activity, fatty acid oxidation, and hepatic gluconeogenesis to spare glucose utilization by peripheral tissues. With the arrival of summer, their metabolism reverts back to an insulin-resistant state. These seasonal metabolic changes are a function of different neurotransmitters in the suprachiasmatic nuclei (SCN)—the mammalian circadian pacemaker—and in the ventromedial hypothalamus (VMH) [21].

During insulin-resistant states, serotonin and noradrenergic levels and activity are increased in VMH, and during insulin-sensitive states these levels decrease or return to normal in animals undergoing seasonal changes in metabolism [23–27]. On the other hand it has been observed that plasma dopamine levels are low in the insulin-resistant state, while with restoration of insulin sensitivity, dopamine levels normalize [28–30].

Alteration of dopaminergic tone in the hypothalamus in the insulin-resistant state leads to modifications in hypothalamic neural activities, which results in excessive production of glucose by the liver, adipose lipolysis, and insulin resistance particularly in the postprandial period. This alteration in hypothalamic dopaminergic tone promotes an increase in hepatic glucose, free fatty acids and triglycerides production, resulting in insulin resistance [28,30–34]. Plasma prolactin concentration peaks during sleep at night in insulin-sensitive individuals whereas the plasma prolactin level is increased twofold during daytime in insulin-resistant humans [35].

Insulin resistance can be induced by destruction of dopaminergic neurons [33]. Bromocriptine administered peripherally or intracerebroventricularly in insulin-resistant animals has been shown to decrease hepatic glucose production, gluconeogenesis, adipose tissue lipolysis, and improve insulin sensitivity [28,31]. Bromocriptine-QR administered within 2 hours of waking reduces elevated prolactin levels during the day restoring dopaminergic activity, which reduces elevated plasma glucose levels [35–39].

**Dosage and pharmacokinetics**

Cycloset is the brand name of a quick-release formulation of bromocriptine. It is available in a 0.8 mg tablet given once a day. A starting dose of 0.8 mg should be increased weekly from 1.6 mg up to 4.8 mg (maximum dose) within 30 days [40].

After oral administration, bromocriptine-QR is rapidly absorbed in minutes [35]. Absorption is much faster on an empty stomach (maximum plasma concentration within 60 min) as compared to fed states. Food delays the absorption and maximum plasma concentration is achieved in 120 minutes. Only 5–10% of the drug reaches the systemic circulation as it...
undergoes extensive hepatic first-pass metabolism. Ninety percent of the drug is excreted through the biliary system [41–43].

Efficacy

The efficacy and safety of bromocriptine in the treatment of T2DM has been evaluated in several studies, including monotherapy trials, two 24-week trials in which bromocriptine was added to sulfonylurea, and in a 52-week randomized, double-blind, placebo-controlled safety trial in which bromocriptine was added to several antidiabetes treatments including insulin (mean baseline HbA1c 8.3%). Significant improvement in HbA1c was reported in patients receiving bromocriptine as an adjunctive therapy to other oral antidiabetic medication [40,44].

A double-blind, placebo-controlled trial enrolled 17 obese subjects with impaired glucose tolerance; patients were randomized to bromocriptine-QR (1.6–2.4 mg d$^{-1}$) daily or placebo for 18 weeks. The treatment group demonstrated a significant reduction in weight (6.3 ± 1.5 kg, p < 0.01) with a decrease in the area under the serum glucose curve of 46% during an oral glucose tolerance test compared to baseline [36].

In a 16-week, double-blind, placebo-controlled study 22 obese subjects with inadequately controlled T2DM were randomized to receive bromocriptine versus placebo. HbA1c was reduced from 8.7 to 8.1%, and fasting plasma glucose reduced from 190 to 172 mg dL$^{-1}$ in the treatment group. In the placebo group HbA1c increased from 8.5 to 9.1%, and fasting plasma glucose increased from 187 to 223 mg dL$^{-1}$ [45].

In a randomized, double-blind placebo-controlled study, addition of once daily bromocriptine-QR to sulfonylurea in patients with inadequately controlled diabetes led to a decrease in HbA1c by 0.55%, and fasting and postprandial glucose by 23 and 26 mg dL$^{-1}$, respectively. The study also reported a decrease in fasting triglycerides by 72 mg dL$^{-1}$ and postprandial by 63 mg dL$^{-1}$ [35].

In another double-blind placebo-controlled clinical trial 40 obese patients with T2DM were randomly assigned to either receive bromocriptine-QR (2.5 mg daily) or placebo for 3 months. Improvement in FPG level was reported in the bromocriptine-QR group. Reduction in HbA1c in the treatment group was from 9.9 ± 0.3 to 9.5 ± 0.2% [46].

Recently a study on efficacy and safety of bromocriptine-QR was carried out in an Indian population. The study enrolled 105 patients and randomized them into three groups; group 1 received only bromocriptine-QR, group 2 received metformin 500 mg twice daily, and group 3 received a combination therapy of bromocriptine-QR plus metformin. The mean reductions in HbA1c at 12 weeks were 0.46%, 0.63%, and 0.74% in groups 1, 2, and 3, respectively. At 12 weeks mean FBS in group 1 was reduced by 16.09 mg dL$^{-1}$ (p < 0.05). In group 2 the decrease was 37.36 mg dL$^{-1}$ (p < 0.05) at 12 weeks, and in the combination group mean FBS was reduced by 44.31 mg dL$^{-1}$ (p < 0.05) [47].

Safety

Side effects with bromocriptine-QR monotherapy are generally mild and transient. The most common side effect with bromocriptine-QR monotherapy is nausea when compared to placebo (26 vs. 5%). Other side effects found in bromocriptine-QR versus placebo were asthenia (15 vs. 8%), dizziness (11 vs. 6%), constipation (11 vs. 4%), and rhinitis (8 vs. 5%) [35].

A large 52-week, randomized, double-blind, multicenter trial of 3095 patients was the first study to assess the overall as well as cardiovascular safety of bromocriptine-QR. Serious side effects were 8.6% versus 9.6% with placebo. Statistically significant reduction in cardiovascular (CV) endpoints defined as a composite of myocardial infarction, stroke, coronary revascularization, and hospitalization for angina or congestive heart failure was reported (bromocriptine vs. placebo 1.8% vs. 3.2%, respectively) [44]. The mechanism of such a reduction in cardiovascular events remains unclear.

Incidence of hypoglycemia with bromocriptine-QR monotherapy is similar to placebo. When added to other antidiabetic agents including insulin, a lack of hypoglycemic adverse effects is observed [35]. Bromocriptine-QR does not increase insulin levels, which seems to be the most likely reason for the lack of increase in incidence of hypoglycemia [39].

Contraindications

Bromocriptine-QR is contraindicated in patients with psychotic disorders as it can exacerbate the underlying disorder, T1DM, and nursing women as it can inhibit lactation. Caution should be observed in patients who are on antihypertensive agents as it can cause orthostatic hypotension. Patients should be assessed for postural hypotension prior to staring treatment with bromocriptine-QR [40,48].

Drug interactions

Bromocriptine-QR is metabolized by the CYPA4 pathway in the liver hence CYPA4 inducers reduce serum concentration of bromocriptine-QR and drugs that inhibit CYP3A4 increase bromocriptine-QR concentration [49].

Caution needs to be observed when medications that are highly protein-bound (e.g. salicylates, sulfonamides, chloramphenicol, and probenecid) are concomitantly used with bromocriptine-QR. Being highly bound to serum proteins, bromocriptine-QR can increase the unbound fraction of these medications. Hence its use may alter the effectiveness of the above-mentioned drugs. Bromocriptine-QR is dopamine agonist and its concomitant use with dopamine receptor antagonists, such as neuroleptics or metoclopramide, can reduce the effectiveness of bromocriptine-QR and vice versa.

When bromocriptine-QR is used with ergot-related drugs prescribed in the treatment of migraines, it can cause a worsening of ergot-related side effects such as nausea, vomiting,
and fatigue. Ergot agents should not be used within 6 hours of bromocriptine-QR [40,48].

Conclusion

Bromocriptine-QR is not only efficacious in improving glycemia but also has a good safety profile. It does not increase the risk of cardiovascular disease and meets the FDA’s new cardiovascular safety guidelines. More studies are needed to elucidate the mechanism of action and biochemical effects of this new, centrally acting medication for treatment of T2DM. Furthermore, longer duration studies should be undertaken to validate cardiovascular safety of bromocriptine-QR. One year may be too short a time to assess cardiovascular safety.

Acknowledgment

The authors wish to thank Allyson Mackay and Philip Walker from Matas Library at Tulane University for assistance in literature research.

References

28. Luo S, Meier AH, Cincotta AH: Bromocriptine reduces obesity, glucose intolerance and extracellular monoamine metabolite


SECTION IX

Management of diabetes: monitoring and other interventions
CHAPTER 51

Implantable pumps and artificial and bio-artificial pancreas system

William V. Tamborlane and Miladys M. Palau-Collazo
Department of Pediatrics, Yale University School of Medicine, New Haven, CT, USA

Key points

• Near-normoglycemia and A1c levels should be achieved and maintained early in the course of the disease, preferably since its diagnosis to decrease the risk of complications. However, too many patients with type 1 diabetes (T1DM) fail to achieve target A1c levels with current open-loop therapy.

• All three components for an external closed loop (CL) artificial pancreas system are already available.

• The feasibility of the CL artificial pancreas has been documented in adults, children, and pregnant women who have demonstrated an increased time spent in target range.

• Given the risk from immunosuppressive therapy from islet cell transplantation outweighs the benefits, the most promising therapy for β-cell replacement in patients with T1DM is a fully automated, CL artificial pancreas.

• Bio-artificial pancreas (BAP) is another promising therapeutic approach to transform the treatment of T1DM.

Introduction

Type 1 diabetes (T1DM) is characterized by a progressive decline in insulin secretion leading to absolute insulin deficiency [1]. The incidence of T1DM is rising with an expected doubling in the number of new cases between 2005 and 2020 in children less than the age of 5 [2]. In the 1980s, major advances were made in the treatment and management of T1DM that initiated the start of the intensive treatment era. The Diabetes Control and Complications Trial (DCCT) was a groundbreaking trial conducted between 1983 and 1993 which conclusively demonstrated that intensive therapy (multiple dose injections of insulin a day or insulin pump therapy) with intensive monitoring compared to standard therapy (two injections of insulin a day) with minimal monitoring was more effective in delaying the onset and slowing the progression of diabetic retinopathy, nephropathy, and neuropathy in patients with T1DM [3]. In the observational study that followed the DCCT, the Epidemiology of Diabetes Interventions and Complications (EDIC) study, the differences in complication development between conventionally treated and intensively treated patients continued to increase despite similar glycemic control during EDIC, a phenomenon referred to as metabolic memory. This observation highlighted the importance of introducing intensive therapy with the aim of glycemic goals as close to normal as possible early in the course of the disease [4]. In order to duplicate the physiologic secretion of insulin by the β cell, major advances in human insulin analogue development, with improved pharmacokinetics, have been achieved coupled with significant advances in delivery systems. In 1964, the idea of developing an artificial pancreas (AP) was conceived by Dr. Arnold Kadish and in 1977, the Biostator, the first experimental closed loop (CL) artificial pancreas, became a reality [5]. This device was very useful in the context of short-term experimental studies but was not suitable for home or portable use due to its large size and the requirement of two intravenous catheters; one to measure blood glucose (BG) and another to infuse insulin and glucose [6,7]. The development of continuous subcutaneous insulin infusion (CSII) with a miniature pump in the 1970s and intraperitoneal insulin (IP) delivery with implantable pumps 20 years later overcame one obstacle to the development of a practically applicable closed-loop system. However, a major obstacle to an artificial pancreas remains, and is being addressed with, the introduction of real-time continuous glucose monitoring devices more than 10 years ago [8,9].

Despite advances in open-loop therapy, most patients with T1DM are still being treated with multiple dose injections (MDI) therapy and most people with T1DM do not meet glycemic goals (A1c < 7%) [10]. Moreover, patients who are able to achieve target BG and A1c levels are at higher risk for severe hypoglycemic events, as well as being frequently exposed to asymptomatic
hypoglycemia, especially during the night [3]. Even under the best of circumstances, the constant demands that patients with T1DM have to implement in order to adequately manage their disease can be overwhelming. Only when the patient can be taken out of the equation and when metabolic control can be maintained with little or no effort, will we have achieved the full potential of recent advances in diabetes technology.

The artificial pancreas

Implanted closed-loop system
An early direction of CL development was based on the assumption that a fully implanted system was the “Holy Grail” of this research and the introduction of the MiniMed MIP 2007 implantable insulin pump was one of the first milestones. This pump is a disk shape infusion system implanted in the abdominal wall with the catheter inserted in the peritoneal cavity. It was refilled every 45 to 90 days with a special insulin preparation of human regular insulin stabilized by a surfactant material at a concentration of 400 IU mL\(^{-1}\) [11–17].

Most of the experience with implantable pumps in open-loop therapy came from the 15 centers that form the EVADIAC association (Évaluation dans le Diabète du Traitement par Implants Actifs) in France. The primary indication for enrollment of patients in implanted pump trials was a failure of intensive insulin treatment using CSII to control hyperglycemia in spite of reinforced patient education and close medical follow-up. Other indications were achievement of target A1c <7% levels at the expense of recurrent severe hypoglycemia and poor diabetes-related quality of life [12,15]. A summary of the early clinical trials of patients with T1DM treated with implantable insulin pumps can be found in Table 51.1.

Despite the potential benefits of implanted pumps, there is no current research being undertaken to develop new and improved implanted pump systems. Insulin under-delivery due to clogging of the pumping mechanism and peritoneal catheter due to insulin aggregation has been a major problem. Localized site problems or pump pocket events (hematomas, skin erosions, pocket infections, pump migrations, local abdominal pain), pump failures (electrical or mechanical failure, premature battery depletion), and clinical complications (ketoacidosis and severe hypoglycemia) are other issues that may require a surgical operation to replace the pump. Even the regular refills of the pump’s reservoir can be challenging [18–22].

It was also believed that optimal performance of a CL artificial pancreas system would require a continuous glucose monitoring system with intravascular implantation of the glucose sensor. While MiniMed reported some encouraging early results with their intravascular sensor that was placed in the superior vena cava, technical and clinical problems have been difficult to overcome [23]. Consequently, no long-term implantable sensor is currently available.

External closed-loop system
In contrast to the difficulties in developing an implantable CL system, all three components for an external system are currently available: transcutaneous continuous glucose sensors that measure interstitial glucose concentrations, external insulin pumps and computer algorithms that determine the amount and rate of insulin delivery based on glucose sensor outputs. In order to achieve the ultimate goal of an outpatient therapy with a fully automated artificial pancreas, the system has to be easily managed by the patient, have safety checks in place to avoid over-delivery of insulin, and be able to respond to the normal daily activities that affect blood sugars levels [10,24].

Table 51.1  Clinical trials of intraperitoneal insulin delivery with implantable insulin pump in patients with type 1 diabetes

<table>
<thead>
<tr>
<th>Year published (Reference)</th>
<th>Number of patients</th>
<th>Length of study (patient-years)</th>
<th>Mean HbA1c (%) ±SD</th>
<th>Episodes of diabetes ketoacidosis</th>
<th>Severe hypoglycemia (episodes/patient-yr)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Initial</td>
<td>Final</td>
<td>Baseline</td>
</tr>
<tr>
<td>1988 [16]</td>
<td>20</td>
<td>18.2</td>
<td>7.6 ± NA</td>
<td>7.0 ± NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(p &lt; 0.05)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1989 [17]</td>
<td>18</td>
<td>28</td>
<td>9.2 ± 0.4</td>
<td>8.2 ± 0.4</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(p &lt; 0.01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1992 [18]</td>
<td>56</td>
<td>73</td>
<td>7.4 ± 1.2</td>
<td>7.1 ± 1.0</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(p &lt; 0.001)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1995 [19]</td>
<td>224</td>
<td>353</td>
<td>7.4 ± 1.8</td>
<td>6.8 ± 1.0</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(p &lt; 0.001)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2002* [20]</td>
<td>59</td>
<td>59</td>
<td>&gt;8.0</td>
<td>7.3 ± 0.8</td>
<td>NA</td>
</tr>
</tbody>
</table>

NA: values not available.

*Done in patients with type 2 diabetes. They reported a 66% reduction in mild and 87% reduction in severe episodes.

\(p < 0.001\)
Continuous glucose monitoring (CGM) devices
Currently available subcutaneous glucose sensors measure interstitial glucose levels using glucose-oxidase based electrochemical methods. The kinetics between glucose in the blood and interstitial fluid are still not fully elucidated but there are well-documented physiologic lag times between changes in glucose levels in the blood and in the interstitium [26]. Intersitial fluid glucose generates an electrical current when it is oxidized in a reaction catalyzed by glucose oxidase [8]. Electrical current data are transmitted to a receiver and then converted into a glucose level that can be measured in mg dL$^{-1}$ or mmol L$^{-1}$ [25,26]. Boyne et al. (2003) found that time differences between changes in blood and interstitial glucose ranged from 4 to 10 min, with the interstitial glucose lagging behind changes in BG levels [26]. An attached transmitter sends the glucose reading every 1 to 5 min to a receiver with a display screen that is either built into an external insulin pump or a separate device [8]. It has alarms and direction arrows that alert the patient about rapid changes in glucose in either direction (rapid drop or increase in blood sugar). The sensors can be used for as long as 7 days and sensor readings need to be calibrated with BG usually twice a day [10].

Hypoglycemia is one of the main factors limiting the management of T1DM and it may be a life-threatening complication of intensive therapy. The DCCT reported a threefold increase in severe hypoglycemia in those treated with intensive therapy [3,27]. The risk of hypoglycemia is higher in children and adolescents, and increases with the duration of illness or with tight diabetes control (hypoglycemia-associated autonomic failure) [28,29]. Postprandial BG levels are also suboptimal in most patients with T1DM and contribute to an elevated HbA1c and wide BG excursions may be implicated in the pathogenesis of diabetes complications [30]. The introduction of continuous glucose monitoring (CGM) systems has allowed clinicians and patients to have a real-time view of their BG and its use has been associated with improvement in glycemic control if the devices are used frequently enough [31–33].

A number of studies have evaluated the efficacy of CGM with current approaches to open-loop insulin delivery. While CGM has not eliminated the risk of severe hypoglycemia, CGM allows patients to reduce elevated HbA1c levels without increasing the risk of severe hypoglycemia [31]. The benefits of CGM are related to the amount of time they are worn by the patient (on average more than 5 days a week) [31,33]. In a multicenter clinical trial funded by the Juvenile Diabetes Research Foundation (JDRF) an improvement in A1c of 0.5% after 6 months of CGM use was seen in those patients above the age of 25 (83% of patients used the CGM 6 or more days a week) [34]. No significant difference in A1c was noted in the younger group (only 30% of patients age 15–24 and 50% of patients age 8–14 used the CGM on average for 6 or more days) [34]. In well-controlled patients, CGM can substantially reduce exposure to biochemical hypoglycemia, while helping to maintain HbA1c levels within the normal range [34].

Although improvements in accuracy and precision have been achieved with CGM devices, they remain difficult to use on a daily basis by most T1DM patients, especially adolescents and young adults. More frequent use of CGM would result if patients were able to gain a greater benefit from wearing these devices.

Insulin pump
Intensive management of T1DM, with the goal of reaching HbA1c targets of less than 7%, can be done with continuous insulin infusion or with multiple daily injections (MDI). In contrast to MDI, where the patient has to mentally calculate the amount of insulin that is needed to cover carbohydrates or correct a high blood sugar, today’s “smart pumps” contain microcomputers that allow preprogramming of different basal rates; calculate the amount of insulin needed for carbohydrate coverage; and calculate correction doses to bring high BG level back in to range and automatically adjust this estimate based on the time since the last correction bolus. They also have the advantage of giving as little as 0.025 units of insulin at a time [35].

The Sensor-Augmented Pump Therapy for A1c Reduction (STAR 3) trial compared the efficacy of Medtronic’s MiniMed Paradigm® REAL-Time system with MDI during a 12-month period [33]. Patients on sensor-augmented insulin pump (SAP) therapy demonstrated a reduction in mean A1c levels that was four times greater than the MDI group (0.8% study vs. 0.2% control (p < 0.001) and the proportion of patients who reached the HbA1c target (<7%) was greater. The rate of severe hypoglycemia did not differ significantly between the groups. In the continuation phase of the STAR 3 trial, MDI subjects were allowed to switch to SAP therapy for 6 months (the crossover group) and SAP subjects were allowed to remain on uninterrupted SAP therapy for a total of 18 months. A1c values were initially lower in the continuing SAP group than in the crossover group (7.4 vs. 8.0%, p < 0.001). There was a statistically significant further A1c reduction in the crossover group after 3 months on the SAP system [32,33]. The improved A1c levels achieved by the SAP group during the first 12 months of the study were maintained at 15 and 18 months. As demonstrated in other studies, sensor wear times of greater than 40% were required to achieve and maintain the metabolic benefits of SAP therapy [33].

Algorithms
Variations of two main control algorithms are currently being used in most of the artificial pancreas clinical trials: the proportional-integral-derivative (PID) algorithm and the model predictive control (MPC) algorithm. The PID algorithm is based on retrospective analysis of prior changes in sensor glucose levels over a fixed period of time (e.g. the last 10 minutes). The insulin delivery calculated on the proportional component is based on how far the actual glucose level is from the desired glucose level; the derivative component varies the rate of insulin delivery based on the rate and direction of change of sensor glucose levels and the integral component is based on
the amount of time BG has remained above or below the target range [10]. PID algorithms essentially look retrospectively to glucose changes over time [24]. The PID algorithm now has an insulin feedback component that takes into account the “insulin-on-board” to avoid late postprandial hypoglycemia that results from delays in absorption of meal-related insulin delivery [24]. MPC algorithms compare the predicted BG with the actual glucose, and then try to predict how much insulin is needed to maintain the blood sugar within a target.

Feasibility studies
The feasibility of the CL artificial pancreas has been documented in adults, children, and pregnant women who have demonstrated an increased time spent in target range. Since CL insulin delivery has reduced but not eliminated problems with hypoglycemia the group from Boston University has developed a bi-hormonal CL artificial pancreas using glucagon as the counterregulatory component.

Even under CL conditions, meal-related BG excursions frequently exceed target levels. This is primarily due to delays in absorption of insulin from the subcutaneous tissue and because the algorithm extends the meal “bolus” over a period of 2–3 hours. The Yale Ultra-Fast Acting Insulin Project was undertaken to explore ways to accelerate the time action profiles of rapid-acting insulin analogues to improve performance of closed and open-loop insulin delivery. Cengiz et al. (2013) have reported that warming the skin around the insulin infusion site to 38.5°C with the InsuPatch (IP38.5) results in an earlier time to peak insulin action (TGLmax) as compared to control glucose clamps without IP activation (No IP) [36]. Weinzimer and colleagues demonstrated that postprandial hyperglycemia can also be ameliorated by “meal announcement”; namely having the patient administer a partial, pre-meal priming bolus. In their study, this hybrid, semiautomatic approach led to an earlier rise in plasma insulin levels and a reduction in peak postprandial glucose levels [37]. Although this approach did decrease postprandial hyperglycemia, it requires manual inputs from the patient, and it is therefore not fully autonomous [37].

Another strategy for reducing post-meal hyperglycemia is by administration of pre-meal injections of pramlintide. Pramlintide is an analogue of the naturally occurring β-cell peptide amylin. It works by delaying gastric emptying, thus slowing carbohydrate appearance, and by lowering meal-stimulated increases in plasma glucagon levels [38]. This mechanism of action is important given that it is now known that there is a dysregulation of α-cell function even during the very early stages of T1DM. Recently, Sherr et al. (2013) demonstrated that glucagon responses to hypoglycemia were lost in many patients with T1DM during the first 2 years of the disease; whereas, during mixed meal tolerance testing (MMTT), increases in plasma glucagon were greater in T1DM than in non-T1DM subjects. These data suggest that early dysregulation of α-cell function contributes to increasing problems with hyper-, as well as hypoglycemia in patients with T1DM [39]. To improve the CL control, Weinzimer and his associates studied eight subjects for 48 h on a CL insulin-delivery system with a PID algorithm with insulin feedback: 24 h on CL control alone (CL) and 24 h on CL control plus 30 mg pre-meal injections of pramlintide (CLP) and no pre-meal manual boluses were given. CLP was associated with overall delayed time to peak BG (2.5 ± 0.9 vs. 1.5 ± 0.5 h; p < 0.0001) and reduced magnitude of glycemic excursion (88 ± 42 vs. 113 ± 32 mg dL⁻¹; p = 0.006) compared with CL alone [38]. The major obstacle to outpatient treatment of T1DM patients is that a number of redundant safeguards must be in place to ensure that no patient is injured by over-delivery of insulin due to a system malfunction. The translation of advances in diabetes technology in CL control into clinical practice will be facilitated by improvement in system hardware and software, some of which are shown in Table 51.2.

While the main danger of CL control is when the system is prompted to increase insulin delivery in response to increasing sensor glucose levels, there are few concerns regarding the safety of shutting down insulin delivery for relatively brief periods of time in response to low sensor glucose levels. The first step in this direction has been achieved with the introduction of the Medtronic’s integrated sensor-augmented pump system, which automatically suspends a pump’s basal insulin infusion for 2 h if sensor glucose levels fall to a preprogrammed threshold level and the patient does not respond to the alarm. In various studies, it has shown to be safe and effective in reducing the amount of time spent in hypoglycemia and reducing nocturnal hypoglycemia [40,41]. Another recent study demonstrated that a 2-hour period of suspension is safe and does not result in ketoacidosis [42].

As previously described, the artificial pancreas has been tested extensively in the hospital in-patient setting under artificial research conditions that do not represent the day-to-day life of our patients. The first outpatient artificial pancreas study used a smartphone computational platform [43] and a number of similar studies have been launched or completed.

Given that currently the risk from immunosuppressive therapy from islet cell transplantation outweighs the benefits, the most promising therapy for β-cell replacement in patients with T1DM is a fully automated, CL artificial pancreas. Given that all three components for the CL are available, the National Institute of Health, the Juvenile Diabetes Research Foundation,
and the industry have constructed partnerships to make this dream a reality.

**Future directions: bio-artificial pancreas**

Bio-artificial pancreas (BAP) systems are devices that contain functional islet cells or bio-engineered β cells. This is an exciting area of research, since development of a bio-artificial pancreas should allow the use of not only allotransplantation but also xenotransplantation of the islets with little or no immunosuppression [44].

The concept of the BAP is to develop an artificial membrane that is permeable to both glucose and insulin and not to the immune system [45,46]. In order to be an acceptable therapy, the BAP should maintain viability and functionality for a prolonged period of time and be easily retrievable or biodegradable [47]. There are different types of membranes and they vary by their geometry and other characteristics. Currently, the three different types of devices that are used are: intravascular, diffusion chambers, and extravascular devices.

The design that most closely mimics the physiologic environment of the islets is the intravascular device. In intravascular devices, islets are encapsulated within one or several hollow biocompatible membrane tubes or fibers and then implanted into the circulatory system via connection to a vascular shunt (artery to vein). Due to direct access to oxygen and nutrients from the blood, islet cell viability and function are maintained and the delay in response to a glucose peak is reduced due to higher diffusion potential [46]. In 2006, Ikeda et al. developed a BAP of this type and were able to achieve normoglycemia in a pancreaticectomized pig [48]. Major problems encountered with this device are surgical implantation and clot formation [46].

In a diffusion chamber BAP module, islets are contained between two semipermeable membranes placed on both sides of a ring-like structure that is implanted in the abdominal cavity. This approach has not been successful. The peritoneal implantation results in poor diffusion between the islets and the surrounding blood. Also, the islets tend to clump with each other and undergo necrosis. Furthermore, the mechanical stress from the small intestine peristalsis can destroy these types of chambers [45]. Edamura et al. (2003) used a diffusion chamber module and performed the xenotransplantation of porcine pancreatic endocrine cells to total pancreactomized dogs. In this experiment, two of the BAPs were completely destroyed, and the remaining one was encapsulated by thin fibrous tissue. On histologic examination on the BAP removed from this dog, necrosis of the pancreatic endocrine cells was observed [49].

Extravascular BAP models can be classified by the size of the device: macrocapsular or microcapsular [46]. Microcapsular devices consist of islets suspended in a polymeric gel material, encapsulated in a biocompatible synthetic membrane. They have been implanted in the peritoneal cavity, the kidney capsule, or subcutaneous tissue. The large surface area allows for substance exchange but once implanted in the body (abdominal cavity), they are not retrievable [44–46]. Living Cell Technologies (LCT) is the first company to enter clinical trials using microspheres containing live porcine islet cells. They developed the product Diabecell® which is implanted into a patient’s abdomen using a laparoscopic approach and patients do not require immunosuppression. Clinical trials began in Russia in 2007 at the Sklifosovsky Research Institute in Moscow. Eight patients with insulin-dependent diabetes received between one and three implants of Diabecell. Having completed a successful phase I/IIa clinical trial in Russia, LCT currently has phase IIb clinical trials in New Zealand and Argentina [45].

Macrocapsular devices can be handled macroscopically. They enclose the islets in a larger hydrogel. The advantages are: smaller surgical implantation risk, easier retrievability, size flexibility, and possibly device re-seeding. Their major drawbacks are limited diffusion of nutrients, slow exchange of glucose and insulin, fibrotic response to the outer wall of the membrane, and loss of physical integrity of the capsule. In animal models, a major reason why BAP has failed is due to the development of a fibrotic and inflammatory response to the chemical surface composition of the device [44–46].

The BAP is a promising therapeutic approach for T1DM. Clinical trials of microencapsulated porcine cells are already on their way [45]. Also, recent advances in pancreatic organogenesis has led to the creation of cells with β-cell-like function which could potentially be used in the development of cellular therapy for the treatment of T1DM.

**References**


17 Saudek CD: Implantable insulin infusion pumps: a case presenta-

18 Selam JL, Micossi P, Dunn FL, Nathan DM: Clinical trial of pro-


20 Duckworth WC, Saudek CD, Giobbe-Hurder A, et al.: The Veter-


27 Battelino T, Phillip M, Bratina N, et al.: Effect of continuous glu-

28 Tsilakian E, Tamborlane W, Xing D, et al.: Blunted counterregu-


30 The relationship of glycemic exposure (HbA1c) to the risk of devel-


33 Bergenstal RM, Tamborlane WV, Ahmann A, et al.: Sensor-


35 Tamborlane WV, Sikes KA: Insulin therapy in children and ado-


40 Coudhary P, Shin J, Wang Y, et al.: Insulin pump therapy with auto-

41 Danne T, Kordonouri O, Holder M, et al.: Prevention of hypo-
glycemia by using low glucose suspend function in sensor-


CHAPTER 52
Pancreas and islet transplantation

R. Paul Robertson
Pacific Northwest Diabetes Research Institute; and the University of Washington, Seattle, WA, USA

Key points
• Pancreas transplantation is very effective in providing normal β- and α-cell function and normoglycemia for type 1 diabetic patients.
• Pancreas transplantation arrests and in some cases reverses complications of diabetes, including kidney, eye, nerve, and macrovascular diseases, and improves quality of life in type 1 diabetic recipients.
• Intrahepatic autoislet transplantation is very effective in providing normal β-cell, but not complete α-cell, function after total pancreatectomy for chronic, painful pancreatitis.
• Total pancreatectomy and autoislet transplantation is very effective in relieving abdominal pain, weight loss, and narcotic usage, and improves quality of life in patients with chronic pancreatitis.
• Alloislet transplantation is somewhat effective in providing normal β-cell, but not complete α-cell, function and a return to normoglycemia with freedom from exogenous insulin.
• Alloislet transplantation is highly effective in preventing hypoglycemia secondary to insulin treatment.

Role of pancreas and islet transplantation in management of diabetes

Pancreas and islet transplantation are normally performed in tertiary care centers that have an active kidney transplant program and that are equipped to handle the complex medical and psychosocial needs of transplant patients. To answer the difficult question of which patient should be recommended for pancreas or islet transplantation, the American Diabetes Association prepared a position statement in 2000 [1]. These recommendations were:

1 Pancreas transplantation should be considered an acceptable therapeutic alternative to continued insulin therapy in diabetic patients with imminent or established end-stage renal disease who have had or plan to have a kidney transplant, because the successful addition of a pancreas does not jeopardize patient survival, may improve kidney survival, and will restore normal glycemia. Such patients also must meet the medical indications and criteria for kidney transplantation and not have excessive surgical risk for the dual transplant procedure. Medicare and other third-party payers of medical care should include coverage for pancreas transplant procedures meeting these criteria. The pancreas transplant may be done simultaneously with, or subsequent to, a kidney transplant. Pancreas graft survival is better when done simultaneously with a kidney transplant.

2 In the absence of indications for kidney transplantation, pancreas transplantation should only be considered a therapy in patients who exhibit these three criteria: 1) A history of frequent, acute, and severe metabolic complications (hypoglycemia, hyperglycemia, ketoacidosis) requiring medical attention; 2) Clinical and emotional problems with exogenous insulin therapy that are so severe as to be incapacitating; 3) Consistent failure of insulin-based management to prevent acute complications. Program guidelines for insuring an objective multidisciplinary evaluation of the patient’s condition and eligibility for transplantation should be established and followed. Third-party payer coverage is appropriate only where such guidelines and procedures exist.

3 Pancreatic islet cell transplants hold significant potential advantages over whole-gland transplants. However, at this time, islet cell transplantation is an experimental procedure, also requiring systemic immunosuppression, and should be performed only within the setting of controlled research studies.

These carefully worded guidelines clearly indicate that pancreas and islet transplantation should not be considered routine care for patients with T1DM. Rather, these procedures are reserved for individuals who are unable to establish control of their glucose levels and/or to have a reasonable quality of life by using conventional nonsurgical measures. The availability of home glucose monitors, multiple preparations of insulin, and new drugs have facilitated the evolution of highly sophisticated therapeutic paradigms so that the majority of type 1 diabetic patients are able to maintain hemoglobin A1c levels that fall within the recommended range. A small number of type 2
diabetic patients have received pancreas transplants, but this is not normally considered to be an option for such patients.

**The pancreas transplantation procedure**

Pancreas transplantation was first begun for the treatment of T1DM in humans in 1966 [2]. Not many procedures were performed up to 1978 because of the low rates of graft survival. Thereafter, important improvements in immunosuppressive therapy developed, most notably the use of anti-T-cell agents and cyclosporine. These new drugs, combined with new surgical techniques and the selection of healthier recipients, led to a much greater use of pancreas transplant for therapeutic intervention.

Typically, an intact pancreas donated by a person who has died is placed into the pelvis of the recipient whose native pancreas is left untouched. Most grafts receive arterial blood from, and return venous blood to, iliac vessels. Less often hepatic portal venous drainage of the allograft is used [3]. Uncommonly, a hemi-pancreas is donated to a sibling by a healthy, living, related donor who has undergone hemipancreatectomy. When a cadaveric organ is used, a small portion of the donor duodenum containing the exit of the pancreatic duct is included and oversewn onto the urinary bladder. In this manner, the exocrine drainage of the donated organ drains into the bladder and is excreted. Alternatively, enteric rather than bladder drainage is used. Advantages of bladder over enteric drainage are the use of urinary amylase to monitor for rejection of the pancreas and the avoidance of small bowel complications such as obstruction and infection. On the other hand, the advantages of enteric over bladder drainage are the avoidance of urinary bladder infection, hematuria, and reflux pancreatitis. In terms of hepatic portal venous drainage, one advantage is the avoidance of hyperinsulinemia that is caused when systemic venous drainage is used [4]. This arrangement eliminates first pass hepatic metabolism of insulin, which normally is approximately 80–90%. It should also be appreciated that because of its ectopic position, it is highly unlikely that the central nervous system establishes physiologically meaningful innervation of the transplanted organ. This is an important consideration since the central nervous system under normal physiologic conditions is considered to be an important regulator of pancreatic islet function.

Transplanted pancreas can undergo both acute and chronic rejection episodes despite the use of immunosuppressive drugs. Management involves hospitalization and intensification of immunosuppression. If the pancreas is transplanted simultaneously with a kidney, detection of rejection is facilitated because an increase in serum creatinine can be easily monitored and used as a signal that both organs are at risk for undergoing a rejection episode. When a pancreas is transplanted with bladder drainage without a kidney, the less sensitive and later indices of decreasing urinary amylase, increasing serum amylase, and increasing blood glucose levels are the only available warnings of rejection. Cystoscopic transduodenal or percutaneous biopsy with ultrasound guidance is used to confirm rejection. A renewed autoimmune attack on the transplanted islet has been suggested as a possible cause of graft demise, but this is probably an unusual event, partly because the recipient receives immunosuppressive drugs that are known to suppress primary autoimmune responses.

Conventionally, long-term maintenance of immunosuppression is provided by triple-drug therapy with either cyclosporine or tacrolimus (both calcineurin inhibitors), azathioprine or mycophenolate mofetil (both antimetabolites), and corticosteroids. Increasingly, more immunosuppressive protocols are excluding corticosteroids because of their severe adverse side effects that greatly compromise quality of life, such as metabolic bone disease and cushingoid appearance. Newer drugs are becoming more increasingly in vogue [5,6].

Serious adverse effects often accompany the surgery required for pancreas transplantation as well as the immunosuppressive drugs used [7]. Perioperative complications leading to laparotomy occur in as many as one of three recipients because of intra-abdominal infections and abscess, vascular graft thrombosis, anastomotic leakage, and duodenal stump leak. Drug-related complications include bacterial and viral infections (particularly cytomegalovirus) and malignancy (particularly skin tumors and lymphoma). Adverse metabolic side effects of the drugs include osteoporosis and insulin resistance as a consequence of using glucocorticoids and decreased renal function as a consequence of using calcineurin inhibitors. Ironically, many of the immunosuppressive drugs, particularly the calcineurin inhibitors and glucocorticoids, have deleterious effects on pancreatic β-cell function. These include decreased β-cell insulin gene expression and insulin content as well as defective insulin secretion in response to glucose [8]. For this reason, newer drugs are being sought that do not adversely affect β-cell function of the transplanted pancreas and do not cause metabolic bone disease.

**The islet transplantation procedure**

Both autoislet and alloislet transplantation usually have used the intrahepatic site for infusion of islets through a catheter. In the autoislet scenario, a patient with chronic, unrelentingly painful pancreatitis undergoes total pancreatectomy for pain relief. The pancreas is used for islet isolation and the islets are returned via the hepatic portal venous system within 2–4 hours while the patient is still in the operating room. Historically, autoislets were not purified; although more recently some centers favor purification to decrease the mass of tissue infused into the portal circulation because of concerns about increased portal blood pressure. Purification, however, considerably decreases the mass of viable islets eventually infused [9]. The alloislet scenario involves patients with T1DM who typically have significant clinical problems with hypoglycemia secondary
to use of exogenous insulin. Success for alloislet procedure is highly dependent on the total mass of islets transplanted. Usually two separate pancreas donors on two separate occasions are needed to attain freedom from exogenous insulin treatment. Another major difference between the two procedures is that autoislet transplantation does not require immunosuppression, whereas alloislet transplantation does. The approach to immuno-suppression is very similar to that used for whole pancreas transplantation. The contrast in the amount of islets required for autoislet and alloislet transplantation is of great interest and points to the likelihood of greater islet damage in the alloislet procedure, possibly because of extensive purification [9], β-cell toxicity of immunosuppressive drugs [8], or recurrence of autoimmune disease. Concern has been expressed about using the liver as the sole transplant site for islets from several points of view [8–12]. The list of complications is substantially different from whole organ transplantation. The problems encountered for autoislet transplants are not intrinsic to intrahepatic islet infusion but rather the surgery involved in total pancreatectomy, which carries the type of risks associated with any major intra-abdominal surgery. On the other hand, alloislet transplantation carries unique risks of bleeding and thrombosis because of the use of a percutaneous metal cannula which is inserted into the liver to advance polyethylene tubing into the hepatic portal venous circulation [13,14]. Very few deaths have been associated with autoislet or alloislet transplantation.

**Success rates for pancreas and islet transplantation**

**Pancreas**

Annual pancreas transplantation rates as well as the success rates for patient and organ survivals have continually increased since 1978. By the year 2011 over 18,000 pancreas transplantations had been performed nationally [15,16] (Figure 52.1). Between the years 1978 and 2011 there were many changes in surgical techniques and immunosuppressive drugs that led to progressively higher patient and graft survival rates. Historically, the majority of pancreases were transplanted simultaneously with kidneys. In 1999, of 1287 pancreas transplants reported, 75% were performed simultaneously with kidneys, 18% were performed after kidney transplantation, and 7% were performed unrelated to previous or concurrent kidney transplantation [3]. The patient survival rate for simultaneous pancreas and kidney transplantation 1 year post-transplant is currently 95% [15,16]. It seems likely that the 5% mortality is due more to the chronic diabetic state than the surgery itself because the vast majority of deaths do not occur in the perioperative period and are usually cardiovascular in nature. In this regard, it is important to bear in mind that the average recipient of a pancreas transplant has had diabetes mellitus for over 20 years prior to the surgery and has had serious complications including macrovascular and microvascular disease. For pancreases transplanted simultaneously with kidneys, the pancreas graft survival rate was 79% at 1 year and 77% at 3 years. Pancreas transplantation after a previous kidney transplantation rates were 69% at 1 year and 59% at 3 years. Pancreas transplanted alone rates were 65% at 1 year and 58% at 3 years. More recently, in the 2009–2011 era, pancreas survival after a simultaneous pancreas and kidney transplantation is 87% at 1 year and 79% at 3 years [16]. For pancreas transplantation alone the rates were 79% at 1 year and 64% at 3 years. The reasons for the greater survival rates for simultaneous and pancreas and kidney transplantation are not clear, although this is the procedure most commonly used, and therefore, the most familiar. Another advantage of the combined procedure is that serum creatinine levels are used to monitor for early kidney rejection, which results in antirejection treatment that may also benefit the transplanted pancreas before manifest islet failure is apparent. When interpreting these data, it is important to remember that none of the patients were assigned to transplant categories on a randomized basis and that the data cited came from multiple transplant centers with varying degrees of experience.

**Autoislets**

Autoislet transplantation was originated at the University of Minnesota where the first procedure was carried out in 1980 [17]. Overall, the recipients have usually been adults, female, and nondiabetic at the time of pancreatectomy. The etiologies have been 41% idiopathic, 17% pancreas divisum, 14% genetic, 9% related to abnormalities involving the sphincter of Odi, and 7% alcoholic [18]. More recently, children with pancreatitis are undergoing this procedure [19]. By the mid-90s it was reported by this group that a minimum of 300,000 islets was required to achieve normoglycemia and insulin-independence in 74% of recipients for >2 years [20]. Many more centers have
more recently accelerated their autoislet transplant programs and are reporting outcomes similar to those published by the University of Minnesota [21–28].

**Alloislets**
The most extensive single center experience with alloislets has been that of the Edmonton group from the University of Alberta. Their initial publication in 2000 demonstrating a 100% success rate in 7 patients receiving islets from an average of two donated pancreases with avoidance of corticosteroid treatment in the immunosuppression regimen [29] sparked great interest. Although the initial successes were largely confirmed from their 2 year post-transplant report, by 5 years enthusiasm for the procedure had diminished because by 15 months post-transplant 50% of the recipients had returned to insulin [30]. The 5-year outcomes revealed a 15% failure rate, a 77% partial success rate, and an 8% complete success rate [30] (Figure 52.2). Subsequent reports from collaborative studies and single centers using the Edmonton protocol have reported data that are similar to that published by the Edmonton group [31–36]. The Collaborative Islet Transplant Registry (CITR) Sixth Annual Report of November 1, 2009 [14] indicates improvement in success rates in achieving insulin independence (70% at 1 year post-transplant, 55% at 2 years, and 36% at 4 years). With increasing frequency publications have appeared from single centers reporting improved islet survival rates and use of singe-donor alloislet grafts [37,38].

**Clinical outcomes on diabetes complications after successful pancreas and alloislet transplantation**

**Beta-cell function**

Pancreas
The major metabolic defect to overcome in patients with T1DM is the total absence of insulin secretion. Their β cells have been killed by the autoimmune process that underlies this disease. Although exogenous insulin therapy has been remarkably effective in maintaining the life of such patients, it is extremely difficult, if not impossible, to maintain 24-hourly normal glucose levels using insulin-based therapy. In contrast, successful pancreas transplantation is remarkably effective in restoring endogenous insulin secretion, independence from exogenous insulin therapy, normal glucose levels, and normal hemoglobin A1c levels [39]. Normalization of fasting blood glucose levels has been documented to exceed several decades. That not only fasting but also postprandial glucose levels are improved can be appreciated when hemoglobin A1c levels in pancreas transplant recipients are compared to those observed in the Diabetes Control and Complications Trial (DCCT). Patients undergoing standard therapy in the DCCT had an average hemoglobin A1c value of approximately 9% whereas those undergoing intensive insulin-based management had an average level of approximately 7%. In a group of successful pancreas transplant recipients followed over 6 years, the average hemoglobin A1c level ranged between 5% and 6% [40].

**Alloislets**
Using a circulating C-peptide level of >0.3 ng mL−1 as indicative of graft function, CITR reports C-peptide positivity in 72% of recipients at 1 year and 41% at 4 years post-transplant [14]. HbA1c levels <6.5% and absence of severe hypoglycemia was reported as 2% pre-transplant, 51–65% at one year post-transplant, and 20–45% at 4 years post-transplant. Restoration of β-cell function by measurements of insulin secretion has been well documented [41,42].

**Alpha-cell function**

Pancreas
Hypoglycemia is one of the major limitations to obtaining optimal hemoglobin A1c levels when using intensive insulin-based management. In nondiabetic individuals, glucagon secretion from the pancreatic α cell occurs promptly when glucose levels fall to approximately 65 mg dL−1. Glucagon released from the α cell immediately enters the hepatic portal circulation through which it travels to the liver to stimulate glycogenolysis. This promptly releases glucose into the hepatic vein and thereafter into the general circulation. Soon after the glucagon response, the adrenal medulla releases epinephrine, which also stimulates hepatic glycogenolysis. Typically, patients who have had T1DM for greater than 5 years have absent glucagon responses, and after 10 years have diminished epinephrine responses, to hypoglycemia. To make matters worse, patients who are recurrently hypoglycemic do not have the usual array of physical symptoms—warmth, sweating, visual disturbances, rapid heart rate, and tremulousness—that accompany hypoglycemia. Because these warning signs are absent, patients often do not realize they are hypoglycemic until blood glucose levels drop as low as 30 mg dL−1, an obviously dangerous situation because
obtundation and even death may follow as values drop even lower. The loss of symptomatic responsiveness to hypoglycemia has been shown to be an example of desensitization since loosening of tight metabolic control with insulin in type 1 diabetic patients is accompanied by a return of symptoms to subsequent hypoglycemia [43].

Successful pancreas transplantation completely restores endogenous glucagon secretion to hypoglycemia (Figure 52.3). Hypoglycemic, hyperinsulinemic clamps, have shown that transplanted patients have virtually identical glucagon responses to those observed in nondiabetic control subjects [44]. Benefits are also observed for epinephrine secretion (Figure 52.4), although this response is not completely restored. As far as the patient is concerned, the most salutary improvement is the complete return of symptom responses during hypoglycemia (Figure 52.4). No differences are found in symptomatology at the various glucose levels developed during a hypoglycemic clamp when comparing transplanted patients to normal controls and both groups have markedly more intense symptoms than nontransplanted type 1 diabetic patients [44].

Alloislets
A peculiarity of transplanting islets into the intrahepatic site is that they fail to secrete glucagon in response to hypoglycemia, although they respond normally to intravenous arginine [45] (Figure 52.5). On the other hand, islets transplanted into nonhepatic sites, such as the intraperitoneal cavity, omentum, serosal surfaces of bowel, and spleen respond to hypoglycemia normally [46]. In essence, this implies that the decreased prevalence of insulin-induced hypoglycemia in alloislet recipients cannot be attributed to improved glucagon secretion and associated improvements in hormonal counterregulation of hypoglycemia. This again raises the question whether improved medical management of insulin-based therapy post-transplant might be playing a role in this favorable outcome of the alloislet transplant. The data in this area also raise the issue whether nonhepatic sites alone or in addition to the hepatic site should be used for alloislet and autoislet islet transplantation to provide glucagon responsiveness to hypoglycemia, something the α cells in the native pancreas of T1DM recipients are unable to do because of the absence of contiguous β-cell counterregulation of α-cell function.

Nephropathy
Pancreas
Renal disease is a major secondary complication of chronic hyperglycemia. Decreased glomerular filtration rate and
mesangial thickening are commonly observed as diabetes progresses. Normal kidneys transplanted into diabetic patients who remain hyperglycemic soon undergo the same deterioration in structure and function experienced by native kidneys. In cases of simultaneous pancreas and kidney transplantation, transplanted kidneys fare much better because the recipients become euglycemic. Abnormal diabetic renal structure in the native kidney undergoes resolution of thickened basement membranes and mesangial accumulation after 10 years of normal glucose levels following successful pancreas transplantation [47] (Figure 52.6).

**Neuropathy**

Motor and sensory neuropathies greatly diminish the quality of life in patients with chronic diabetes. Sensory neuropathy can be expressed as chronic numbness of the skin, particularly in the feet, and can also be expressed as chronic pain and exquisite sensitivity to touch. After several years of successful pancreas transplantation and return to euglycemia, indices of sensory and motor nerve function are stabilized. It has been demonstrated that parameters of nerve function including physical exam, motor and sensory nerve conduction, and cardiorespiratory reflexes can be stabilized for up to 10 years in successful recipients [49]. Even more impressively, the marked diminution in life expectancy that accompanies autonomic dysfunction in chronic diabetes, which approximates 50% at 5 years, is greatly improved. In one study, a group of type 1 diabetic patients with autonomic insufficiency who had successful pancreas transplantation and return to euglycemia had a 20% mortality at 5 years versus an approximate 70% mortality in failed pancreas transplant recipients [50].

**Retinopathy**

Pancreas

The impact of successful pancreas transplantation on retinopathy has been equivocal. Several studies comparing transplanted to nontransplanted patients with chronic diabetes have indicated no significant differences in abnormal retinal changes. This generally gloomy conclusion was more recently challenged by a report indicating that less laser therapy for retinopathy was required in patients undergoing pancreas and kidney transplantation compared to those undergoing kidney transplantation alone [51].
Alloislets
As yet, there have been no large series of studies of the effect of successful alloislet transplants on diabetic retinopathy. Preliminary reports of beneficial effects on eyes have appeared, as noted by Fiorina et al. [47].

Macrovascular disease
Initially, there was no indication that macrovascular disease in patients with diabetes may be improved by a return to hypoglycemia following pancreas transplantation. Concern existed that the hyperinsulinism that is caused by systemic venous drainage of the transplanted pancreas may be harmful in terms of accelerating atherosclerosis. However, if anything, circulating lipids are under better control following pancreas transplantation and there have been no reports of accelerated atherosclerosis in transplant recipients. To the contrary, three fairly recent reports suggest that successful pancreas transplant recipients have improved vascular status. Two studies involved examination of intima media thickness of the carotid artery. In one report, thickness was significantly less in patients who had undergone pancreas and kidney transplantation compared to patients undergoing kidney transplantation alone [52]. In the second study, carotid intima media thickness 4 years after successful pancreas transplantation was no different than that found in a nondiabetic control group, but was less than that found in a pre-pancreas transplant diabetic group [53]. However, both studies were cross-sectional and did not obtain paired measures in the same patients before and after pancreas transplantation. On the other hand, a study of coronary artery patency indicated less narrowing over time in a group of successful pancreas transplant patients studied prospectively compared to a group of patients who failed transplantation [54].

Alloislets
As yet, there have been no large series of studies of the effect of successful alloislet transplants on macrovascular disease. Preliminary reports of beneficial effects on vessels have appeared, as noted by Fiorina et al. [47].

Quality of life
Studies consistently demonstrate benefits for successful recipients of pancreas and auto- and alloislet transplantation. Although this outcome is the most clinically difficult to objectively measure, it is clearly the most important indicator of success or failure.

Pancreas
Virtually all published studies report that the quality of life after pancreas transplantation improves regardless of whether a kidney transplant is involved or not [55]. The most common question asked is whether the recipients would prefer a successfully transplanted organ and the risks of chronic immunosuppression or a return to their original diabetic state and insulin-based therapy with the hazards of hypoglycemia. Patients usually report they prefer taking immunosuppressive drugs as the price of a functioning pancreas, normal glucose levels, and decreased risk of developing secondary complications of diabetes, including hypoglycemia. An important caveat is that most diabetic patients who receive transplants have had serious difficulty avoiding extremely high and low glucose levels. Clearly, the worse the quality of life before transplantation, the more likely it is to improve after successful surgery.

Autoislets
Relief from pain in autoislet recipients (improvement or complete resolution) is reported by 80–86% of recipients at 6 months to 3 years after surgery and half have withdrawn completely from narcotics used pre-operatively [18]. Quality of life has generally improved, although not infrequently patients continue to have bouts of abdominal pain that are different in nature than the pain they associated with chronic pancreatitis. This pain is usually ascribed to temporary episodes of partial bowel obstruction secondary to intra-abdominal adhesions. One might expect recipients of intrahepatic autoislets to have difficulties with hypoglycemia because some use long-acting insulin and they have only partially functional intrahepatic α cells. However, the scientific literature is silent in this regard.

Alloislets
The main indication for alloislet transplantation is recurrent hypoglycemia secondary to exogenous insulin treatment. Using this measure, alloislet transplantation has been reported to be very successful [56,57]. Severe hypoglycemic events were reported by the 2009 CITR report [14] as 81% pre-transplant, and <10% by 4 years post-transplant. CITR also notes that even after complete graft failure there is enduring protection against severe hypoglycemia. This phenomenon has not yet been explained but raises the possibility that more sophisticated insulin-based management associated with transplant centers, independent of the success of alloislet grafts, contributes significantly to the lesser degrees of hypoglycemia encountered post-transplant compared to the pre-transplant situation.

References

15 Gruessner AC, Sutherland DE: Pancreas transplant outcomes for United States (US) cases reported to the United Network for Organ Sharing (UNOS) and non-US cases reported to the International Pancreas Transplant Registry (IPTR) as of October, 2000. Clinical Transplantation 2000;45–72.


**CHAPTER 53**

**Hypoglycemia and other complications of insulin therapy**

Stephanie A. Amiel  
King's College London and King's College Hospital, London, UK

---

**Key points**

- Hypoglycemia is common in insulin therapy for type 1 and late (insulin-deficient) type 2 diabetes.
- It results from excess insulin action plus defects in the body's usual counterregulatory mechanisms.
- Loss of subjective awareness of hypoglycemia increases the risk of severe episodes, with loss of cognitive function and ability to self-treat, 6- to 17-fold.
- Exposure to hypoglycemia induces defects in the normal defenses against hypoglycemia which can be restored by hypoglycemia avoidance.
- Fear of hypoglycemia is a major barrier to achieving good glycemic control and quality of life.
- Weight gain with insulin therapy is multifactorial but can be a barrier to patient compliance with therapy.
- Other complications of insulin therapy include local reactions at injection sites and rare cases of local or systemic insulin allergy.

**Introduction**

Hypoglycemia—low blood glucose—is the most important acute complication of glucose-lowering therapies. Hypoglycemia—and fear of hypoglycemia—limits attempts to achieve normoglycemia by pharmacologic therapies in people with diabetes and thus limits the ability to reduce the incidence of both micro- and macrovascular complications through normalized blood glucose control. In health, hypoglycemia only occurs in extreme circumstances, such as after extreme sport. In diabetes, hypoglycemia is common but its true prevalence varies widely, according to how it is defined and the clinical circumstances of the person's diabetes and its management.

Other associations of insulin therapy include failure to achieve optimized glucose control, weight gain, problems at the site of injection, rare local and very rare systemic insulin allergy. The list is short, as exogenous insulin is a (fairly) natural product and insulin therapy is effectively replacement of the missing hormone.

**Definition of hypoglycemia**

Surprisingly, definition of what constitutes hypoglycemia remains controversial. Hypoglycemia may be defined clinically, in terms of a syndrome associated with the low blood glucose concentration, or biochemically, as a quantitative blood (or plasma) glucose concentration below normal.

The American Diabetes Association (ADA) recently reviewed its descriptions of hypoglycemia [1]. Clinically, it describes five hypoglycemia syndromes, based on the severity of the clinical picture. Thus *severe hypoglycemia* is defined as that requiring assistance of another person to treat the hypoglycemia, implying that the plasma glucose has fallen too low to sustain normal cortical (intellectual) function, and the person is too incapacitated to recognize the situation and/or treat it themselves by eating. No defining glucose concentration is given. This definition may be too loose for young children, who will always need help treating a hypoglycemic event, and subcategorization into episodes involving unconsciousness, coma or seizure, and episodes requiring parenteral therapy (intramuscular glucagon or intravenous glucose) is added. Some authorities, particularly those working with children, require one or other or both of these events to be present to define severe hypoglycemia (e.g. [2]). Severe hypoglycemia is thought to occur at rates of 115 to 320 per 100 patient-years in adults with type 1, and 35 to 70 episodes per 100 patient-years in T2DM [3,4].

*Mild hypoglycemia* describes episodes in which the patient perceives symptoms compatible with hypoglycemia (see later) and self-treats by eating, associated with a measured low
plasma or blood glucose concentration. In the ADA document, this category is missing, effectively replaced by documented symptomatic hypoglycemia, which is described as the experience of "typical" hypoglycemia symptoms, in the presence of a measured low glucose concentration, the definition of which is discussed later. Self-management is implied, as need for assistance results in a definition of severe hypoglycemia. Some authorities also recognize moderate hypoglycemia in which the patient is aware and self-treats but the episode includes significant disruption to current activity. This can be useful in quantitating hypoglycemia experience, for example in research studies, but is necessarily subjective. Asymptomatic hypoglycemia is an episode in which plasma glucose is measured as low but for which the patient has no subjective awareness. Clinically, this is diagnosed when a patient is testing his or her blood glucose coincidentally, for example before a meal or an insulin dose, and finds it to be unexpectedly below normal. Asymptomatic hypoglycemia is also diagnosed when someone other than the patient first notices the hypoglycemia, as signs of hypoglycemia are manifested in the absence of any subjective awareness. Regular experience of asymptomatic hypoglycemia is described as impaired awareness of hypoglycemia [5]. It affects about 20% of people with T1DM of more than 2 years' duration [6] and probably about 10% of people with late, insulin-requiring T2DM [7]. It is important to diagnose, as it increases risk of a severe hypoglycemia sixfold in T1, and insulin-requiring T2DM [7]. It is important to diagnose, as it increases risk of a severe hypoglycemia sixfold in T1, and insulin-requiring T2DM [7].

The ADA goes on to define probable hypoglycemia as episodes in which a patient experiences typical symptoms of hypoglycemia and recovers on self-administering glucose, but for which no confirmatory blood test is available. Pseudohypoglycemia is used to describe episodes in which a patient experiences hypoglycemic symptoms but a measured blood test shows unequivocal absence of hypoglycemia, with the plasma glucose concentration either in or above the normal range. The plasma glucose each individual responds to is greatly influenced by that person's recent glycemic experience and pseudohypoglycemia is typically seen in people accustomed to running high blood glucose concentrations and in "poor" diabetic control. This phenomenon has in the past been called relative hypoglycemia, which is perhaps less pejorative.

The aforementioned descriptions are generally accepted and differences in emphases with other classifications are minor. The biochemical definition of hypoglycemia is, however, more controversial. Many textbooks define the range of normal plasma glucose as between 3.5 and 7 mmol L\(^{-1}\) (63–126 mg dL\(^{-1}\)). This would allow us to define hypoglycemia as a plasma glucose of less than 3.5 mmol L\(^{-1}\) (63 mg dL\(^{-1}\)), a definition that has considerable clinical utility. During experimentally induced hypoglycemia, endogenous insulin secretion is reduced and glucagon release occurs at arterialized plasma glucose concentrations of around 4 mmol L\(^{-1}\) (72 mg dL\(^{-1}\)) in healthy people, and because these responses may be diminished in a second experimental hypoglycemia after exposure to this value [8], the earlier referenced American publication defines a plasma glucose of 3.9 mmol L\(^{-1}\) (70 mg dL\(^{-1}\)) as hypoglycemia, making the occurrence of biochemical hypoglycemia much more common and limiting its utility as a predictor of severe hypoglycemia [9]. In people with insulin-deficient diabetes, the counterregulatory (protective) response to hypoglycemia does not include either reduction in circulating insulin or increase in pancreatic glucagon release, so the clinical relevance of a definition of hypoglycemia at 3.9 mmol L\(^{-1}\) is doubtful. Most programs training patients in insulin use set a lower limit of the desirable glucose targets the patient should be trying to achieve as >3.9 mmol L\(^{-1}\) (usually 4 mmol L\(^{-1}\) or higher), so defining ≤3.9 mmol L\(^{-1}\) as an indication to eat to avoid hypoglycemia is universally accepted. Indeed the ADA document refers to this value as an "alert value" while continuing to espouse its use as a definition of hypoglycemia also [1]. Others prefer the clinically unequivocal value of <3.5 mmol L\(^{-1}\). This definition is supported for children, in a guideline written under the umbrella of the International Society for Pediatric and Adolescent Diabetes (ISPAD) in 2009. This describes plasma glucose concentrations between 3.3 and 3.9 mmol L\(^{-1}\) (59 and 70 mg dL\(^{-1}\)) as placing a child at risk of severe hypoglycemia and supports as usual use a cut-off for defining hypoglycemia as <3.6 mmol L\(^{-1}\) (65 mg dL\(^{-1}\)), adding a rider that 3.9 mmol L\(^{-1}\) should be used as the lowest end of any desirable target range to keep in line with the ADA guidance [10]. This advice reinforces the clinical importance of separating out true hypoglycemia, and setting a lower limit for an optimal target range, which should be higher.

Although severe hypoglycemia is almost invariably defined clinically, a plasma glucose value of <3 mmol L\(^{-1}\) (54 mg dL\(^{-1}\)) (the arterialized plasma glucose concentration at which cognitive function is detectable [11]), is sometimes used to define clinically important hypoglycemia, and failure to perceive symptoms when a home blood glucose test is below 3 mmol L\(^{-1}\) is sometimes taken as an indicator of impaired awareness [12]. This definition avoids over-diagnosing impaired awareness of hypoglycemia, which has impact on people's personal freedoms (see later). In children, it has been recommended to define asymptomatic hypoglycemia if the child has no symptoms under 3.6 mmol L\(^{-1}\) (65 gm dL\(^{-1}\)) [10].

Definitions are further complicated by the fact that plasma glucose concentrations are, on average, 10% higher than whole blood glucose concentrations, as red cells occupy space, and venous blood will have a lower glucose concentration than arterial, because glucose will have been removed by metabolically active tissue. Home blood glucose monitoring uses finger prick blood glucose but meters are generally calibrated to plasma glucose values. The error range on many home blood glucose monitoring systems may also be large. New guidelines from the International Organization for Standardization (ISO) will require 99% of readings in the range 4.2 mmol L\(^{-1}\)
(75 mg dL$^{-1}$) or higher to be within 15% of the reference range, and readings less than this value should not differ from the reference range by more than 0.83 mmol L$^{-1}$ (15 mg dL$^{-1}$) [13]. Many currently commercially available systems struggle to reach these constraints, even when quality control is optimized [14].

Using interstitial glucose measurement to diagnose hypoglycemia is also not simple. “Continuous glucose monitoring” (CGM) from sensors placed in the subcutaneous tissue is calibrated to finger-prick glucose measurements, which are generally calibrated to plasma glucose equivalents. The CGM readings may lag behind or run ahead of plasma glucose, especially when its concentration is changing rapidly, by as much as 20 min [15]. For robust detection of clinically important hypoglycemia, CGM values of <3 mmol L$^{-1}$, or even <2.2 mmol L$^{-1}$ have been used [3]. In line with the earlier point about the difference between diagnosing true hypoglycemia and detecting a glucose concentration at which action should be taken to prevent hypoglycemia, care should be taken to set targets for alarms for hypoglycemia when using real-time CGM, to strike the right balance between too many false positives and achieving hypoglycemia protection. While the literature has suggested using 4 mmol L$^{-1}$ as an alarm threshold for hypoglycemia, our clinical practice is to start at 3.5 mmol L$^{-1}$, adjusting the value according to patient characteristics.

**Physiologic responses to hypoglycemia**

The brain requires a constant supply of oxygen and glucose, depending on a well-sustained blood flow with an adequate plasma glucose concentration. Equally, high plasma glucose has pathologic effects. The body has a sophisticated system of glucose control which is deranged in diabetes (Figure 53.1 (a)).

Circulating glucose concentrations are determined by the rate of glucose entering the circulation from endogenous stores in the liver and to some extent kidney, and by consumption and digestion of carbohydrate in eating and drinking, balanced by the rate at which glucose is removed from the circulation by metabolically active tissue. Control of energy balance is complex, beyond the scope of this chapter, but much of it occurs centrally, in the brain, which detects a falling glucose supply by changes in activation of glucose-sensing neurons, concentrated in brain regions such as the ventromedial hypothalamus (reviewed in [16]). The response to a threatened glucose lack is a stress response but includes first a pancreatic response with reduced insulin secretion and enhanced glucagon release, de-repressing and stimulating glycolysis and endogenous glucose production to arrest the plasma glucose fall. In studies where experimental hypoglycemia has been induced by insulin infusion (often with glucose infusion to control the hypoglycemia) this happens at an arterialized plasma glucose of around 3.8 mmol L$^{-1}$ (68 mg dL$^{-1}$) [17]. At a slightly lower glucose, the sympathetic nervous system is activated, adrenaline and noradrenaline concentrations rise in the circulation, further enhancing endogenous glucose production through glycolysis, supporting also gluconeogenesis, in part through enhanced supply of suitable precursors via lipolysis, associated with reduced rate of glucose uptake by peripheral tissues such as muscle and fat. Heart rate and stroke volume increase, and gastric emptying accelerates, with the effect of enhancing tissue and brain glucose supply. Should the glucose continue to fall, driven by continued availability of insulin, cortisol and growth hormone are released further to support gluconeogenesis and reduce peripheral glucose uptake. Clinical episodes of hypoglycemia result from excess insulin action, plus a failure of these counterregulatory mechanisms.

Symptoms of hypoglycemia are in part associated with the stress response. Sometimes termed autonomic, neurogenic, or adrenergic, they include sweating, palpitation, shaking, and hunger [18]. Associated signs include sweating, pallor, tremor,
tachycardia (or very occasionally bradycardia and arrhythmia), widening of the pulse pressure. Another group of symptoms are associated with reduced glucose supply to the brain and are termed neuroglycopenic. These include feeling drowsy, inability to concentrate, feeling confused, having difficulty speaking and in coordination. Associated signs are slurred speech, inappropriate behavior or emotional responses. When symptoms of hypoglycemia are classified by statistical means, three groups of symptoms are identified. In addition to those classically fitting the descriptions of autonomic and neuroglycopenic, a third group of symptoms including blurred vision, headache, and malaise are found [18]. Although neuroglycopenic symptoms probably occur at lower glucose concentrations than adrenergic symptoms, the important point clinically is that it is the patient who feels them. While the patient remains able to experience a symptom, recognize it as associated with hypoglycemia and respond by ingesting carbohydrate, the hypoglycemia is symptomatic and not severe. This remains true even though the patient may describe the episode as very severe, in the sense that he or she experienced very unpleasant and severe symptoms.

In children and the elderly, the clinical appearance of hypoglycemic events is different. In children, behavioral signs are common, and in young children may be the main indication [19]. Such behavioral signs include irritability, erratic behavior, nightmares, and inconsolable crying [10]. In the elderly, at least in people with T2DM, a group of neurologic symptoms associated with coordination and articulation are described [20].

**Defects in counterregulation associated with diabetes treatments**

The treatment of insulin-deficient diabetes is artificial elevation of circulating insulin concentrations by subcutaneous administration of exogenous insulin, which reverses the gradient between the concentration in the hepatic portal vein and that in the peripheral circulation. A degree of peripheral hyperinsulinemia is required to approach normal hepatic insulinization, which may enhance hypoglycemia risk. When insulin secretagogues (sulfonylureas, meglitinides) are used, this is less of a problem but with both forms of therapy the important point is that the circulating insulin concentrations are no longer under endogenous control and remain elevated even if plasma glucose is falling (Figure 53.1(b)). It has also been shown that people with insulin-deficient diabetes do not mount a glucagon response to hypoglycemia [21]. This appears at least in part to be because of loss of an as yet unidentified signal between the β cell and the α cell—if insulin secretion is maintained during hypoglycemia, for example by sulfonylurea induction of the hypoglycemia, the α cell is not responsive to the hypoglycemia [22]. People with insulin-deficient T2DM show a similar defect in glucagon responses to hypoglycemia as described in T1DM [23]. If exogenous insulin is infused via an intraperitoneal route (when about half will be absorbed into the portal circulation), glucagon responsiveness to hypoglycemia may be better maintained.

Loss of both insulin and glucagon responses to hypoglycemia renders the diabetic person dependent on the rest of the stress response and most importantly the symptomatic response as defense against hypoglycemia profound enough to cause cognitive impairment and confusion (Figure 53.1(b)). It is clinically important—loss of C-peptide, indicating more profound loss of endogenous insulin, is an important risk factor for severe hypoglycemia and even imperfect insulin secretory function after islet transplantation provides protection against severe hypoglycemia [24,25]. In the Diabetes Control and Complications Trial in people with T1DM randomized to intensive therapy, absence of measurable C-peptide before or during the first 6 years of the study was associated with higher rates of severe hypoglycemia, illustrating the importance either of C-peptide itself in defense against hypoglycemia by some mechanisms as yet unknown or as an indicator of more complete loss of endogenous insulin with no ability to reduce hepatic exposure to insulin as circulating glucose concentrations fall, and an associated loss of glucagon responses [23]. While there are many data supporting the link between loss of C-peptide and severe hypoglycemia risk, one recent small study has described defective glucagon responses in young people newly diagnosed with T1DM who retained C-peptide responses to a mixed meal, so the impaired glucagon responses may be multifactorial [26]. With increasing diabetes duration, further defects develop, with evidence for impairment of catecholamine responses to hypoglycemia in people with long duration diabetes (over 15 years) [21]. However, about 75% of people retain subjective awareness of hypoglycemia, highlighting that subjective awareness is associated with but not wholly dependent upon the stress hormone responses [27].

Associated with the stress hormone response there are other acute responses to a hypoglycemic episode. Metabolic responses include a fall in plasma potassium. There is an inflammatory response, with release of inflammatory cytokines and circulating clotting factors increase [28]. As described earlier, heart rate and stroke volume increase; brain blood flow rises; gastric emptying accelerates.

Hypoglycemia during sleep at night may not be associated with any counterregulatory response [29].

**Impaired awareness of hypoglycemia**

As reported earlier, impaired hypoglycemia is associated with a 6- to 17-fold increase in severe hypoglycemia. Loss of awareness of hypoglycemia affects approximately 20% of adults with T1DM of more than 2 years’ duration [6], and is likely to increase in prevalence with increasing duration [27]. It has been described in children, with just over 20% of one sample identified as having impaired awareness, robustly identified but including occurrence of severe hypoglycemia in the definition.
In an audit of a UK program of structured education in flexible insulin therapy for adults with T1DM, with a mean duration of diabetes of 18 years, 40% reported not feeling symptoms of hypoglycemia until their glucose was under 3 mmol L\(^{-1}\) or not perceiving symptoms at all [12]. This probably reflects both the long duration of diabetes in people coming for the training program and a preferential referral of people with problematic hypoglycemia. Parenthetically, 1 year after training in flexible insulin therapy, 43% of those reporting impaired awareness at baseline had recovered awareness, although a small number of people also reported new reduced awareness. Overall the rate of impaired awareness fell from 40 to 33% [12].

Loss of subjective awareness of hypoglycemia is associated with a failure of the counterregulatory responses to hypoglycemia, as measured in an experimental setting (Figure 53.1(c)). People with impaired hypoglycemia awareness failed to arrest the glucose fall in response to low-dose insulin infusion, and this failure, and the lack of symptoms, was not associated with classical diabetic autonomic neuropathy [30]. Failure to counterregulate in response to low-dose insulin infusion was found to predict risk of severe hypoglycemia during subsequent intensification of insulin therapy [31]. Disappointingly, intensified therapy did not restore defective glucagon responses to hypoglycemia and indeed impaired adrenaline responses seemed to be induced by the tightening of glycemic control [32]. Clinically, it had been noted that people using the then new insulin pump therapy were “tolerating” hypoglycemia better than when using conventional injection therapy with looser diabetes control.

Using intensified insulin therapy as a model, loss of subjective awareness of hypoglycemia was shown to be associated not only with inability adequately to arrest a glucose fall during low-dose insulin infusion [33], but also to be associated with impaired counterregulatory hormone responses to a fixed stepped hypoglycemic challenge [34]. Catecholamine responses and symptoms occurred at lower glucose concentrations and were diminished at any given glucose concentration in people using intensified insulin therapy (Figure 53.2). The patients were also experiencing more frequent hypoglycemia in daily life. Similar defects in counterregulation were later shown to be inducible in hypoglycemia naïve healthy volunteers and in people with T1DM by antecedent exposure to hypoglycemia [35]. Surprisingly, quite modest exposure was required to induce the defect—2 hours at 3 mmol L\(^{-1}\) in the afternoon following a controlled stepped reduction in arterialized plasma glucose to 2.8 mmol L\(^{-1}\) was enough significantly to impair adrenaline, glucagon, cortisol, and symptomatic responses to a repeated stepped hypoglycemic challenge the next day. Later studies found that 30 min at 3 mmol L\(^{-1}\) was enough to reduce symptomatic responses [36]. Similar data have been found in people with diabetes. *Nocturnal hypoglycemia can create defective awareness of hypoglycemia the next day.*

Experimentally, impaired hypoglycemia awareness, and in most studies, counterregulatory responses, can be restored by careful avoidance of hypoglycemia in daily life [37]. In studies done in the time before online glucose sensing, hypoglycemia of <3 mmol L\(^{-1}\) on intermittent home blood glucose monitoring was avoided for at least 3 weeks to restore responses not just in people with long duration tightly controlled T1DM and impaired hypoglycemia awareness but also in people with poorly controlled T1DM and intermittent severe hypoglycemia. In practice, such avoidance can be difficult for patients to maintain (Figure 53.3) (see restoration of awareness).

The brain’s responses to hypoglycemia

Glucose sensing systems are found in the brain and in the hepatic portal vein, monitoring glucose supply to the brain and glucose entry from the gastrointestinal tract. In preclinical studies, prolonged exposure to hypoglycemia was associated with an upregulation of brain glucose uptake [38,39]. In human volunteers, brain glucose uptake measured by the Kety-Schmidt arteriovenous difference technique was increased during hypoglycemia after 58 hours’ exposure to circulating glucose concentrations of 2.9 mmol L\(^{-1}\) or 52 mg dL\(^{-1}\) (except after meals) and in people with intensively treated T1DM [40,41]. The hypothesis was that hypoglycemia induced an upregulation of brain glucose transport such that brain glucose concentrations might be better maintained in subsequent episodes, so the hypoglycemia was not noticed and no stress response mounted. Yet clinically, onset of cognitive dysfunction in hypoglycemia occurs at glucose...
concentrations no different from or only slightly lower than usual in people with delayed onset of counterregulation, suggesting that any similar mechanism in hypoglycemia exposure in insulin therapy in man is not global. Human neuroimaging studies have failed to show altered glucose uptake in people exposed to different blood glucose concentrations (e.g., [42, 43]).

Rodent studies showed networks of neurons in the basal ganglia and hypothalamus activated by glucoprivation. Lowering glucose in the ventromedial hypothalamus initiated a hyperglycemic stress response, with little response to systemic hypoglycemia if glucose concentrations were maintained there. In human neuroimaging studies, changes in perfusion of the hypothalamus have been demonstrated as plasma glucose was lowered to 4.3 mmol L$^{-1}$ (77.2 +/− 2 mg dL$^{-1}$), prior to the measurable release of catecholamines. In healthy people, an evolution of regional brain activation has been described, involving thalamus, anterior cingulate cortex, the hypothalamic pituitary axis, and prefrontal regions (Figure 53.4) [44]. When comparing people with T1DM with and without impaired awareness of hypoglycemia, reduced activation of brain regions involved in interoception (perception of the body’s internal state); sympathetic responses and stress responses is seen, but also altered responses in prefrontal brain regions involved in reward pathways and hedonic perception [45].

Various experimental protocols have been used to investigate further the pathogenesis of impaired awareness of hypoglycemia, seeking evidence to inform a treatment that might restore the impaired defenses against subsequent hypoglycemic
stimuli. Available evidence implicates dopaminergic pathways and opiate receptors [16]. It is hypothesized that brain glucose sensors use similar intracellular mechanisms to those of the pancreatic β cell to respond to changes in available glucose, with a fall in plasma glucose driving release of an inhibitory neurotransmitter to de-repress the hyperglycemic stress responses, and there has been interest in seeing whether reagents that alter insulin secretion or opiate receptor pathways may also affect counterregulatory responses to hypoglycemia [16].

**Risk for severe hypoglycemia**

Apart from defective counterregulatory hormone responses associated with long duration diabetes and prior exposure to hypoglycemia, risk of hypoglycemia is increased by other factors that impair the counterregulatory response (Table 53.1). These include lack of endogenous glucose stores as may occur

<table>
<thead>
<tr>
<th>Table 53.1 Factors increasing risk of hypoglycemia in insulin-treated diabetes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Non-remidal patient characteristics</strong></td>
</tr>
<tr>
<td>Age (under 5 or elderly)</td>
</tr>
<tr>
<td>Diabetes duration</td>
</tr>
<tr>
<td>Female gender</td>
</tr>
<tr>
<td>C-peptide negativity</td>
</tr>
<tr>
<td>Comorbidities and polypharmacy</td>
</tr>
<tr>
<td>Enhanced insulin sensitivity of early pregnancy</td>
</tr>
<tr>
<td><strong>Precipitants of individual events</strong></td>
</tr>
<tr>
<td>Insulin dose error</td>
</tr>
<tr>
<td>Lower than anticipated carbohydrate intake</td>
</tr>
<tr>
<td>Exercise</td>
</tr>
<tr>
<td>Alcohol excess</td>
</tr>
<tr>
<td>Massage or heating of injection site</td>
</tr>
<tr>
<td>Co-incident use of recreational drugs</td>
</tr>
<tr>
<td><strong>Prolongation of insulin action/reduced insulin clearance</strong></td>
</tr>
<tr>
<td>Exogenous insulin therapy</td>
</tr>
<tr>
<td>Insulin secretagogue therapy</td>
</tr>
<tr>
<td>Abnormal liver function</td>
</tr>
<tr>
<td>Renal failure</td>
</tr>
<tr>
<td>Hypothyroidism</td>
</tr>
<tr>
<td>High insulin-binding antibodies</td>
</tr>
<tr>
<td><strong>Impaired endogenous glucose production</strong></td>
</tr>
<tr>
<td>Liver failure</td>
</tr>
<tr>
<td>Alcohol toxicity</td>
</tr>
<tr>
<td>Glycogen storage disease</td>
</tr>
<tr>
<td>Malnutrition</td>
</tr>
<tr>
<td><strong>Loss of counterregulatory function</strong></td>
</tr>
<tr>
<td>Impaired awareness of hypoglycemia</td>
</tr>
<tr>
<td>Hypopituitarism</td>
</tr>
<tr>
<td>Addison’s disease</td>
</tr>
<tr>
<td>Growth hormone deficiency</td>
</tr>
<tr>
<td><strong>Failure of glucose absorption</strong></td>
</tr>
<tr>
<td>Celiac disease</td>
</tr>
<tr>
<td>Eating disorder</td>
</tr>
<tr>
<td>Exocrine pancreatic failure</td>
</tr>
<tr>
<td>Other malabsorptive disease</td>
</tr>
</tbody>
</table>

*Hypoglycemia may be delayed, commonly occurring in the early hours of the next morning or after breakfast next day.*

in hepatic failure, suppression of hepatic glucose production by alcohol, reduced glycogen stores as in malnutrition, and rare glycogen storage diseases; impaired insulin clearance as in liver or renal failure, hypothyroidism or high levels of insulin-binding antibodies (less common in the present era of purified insulin preparations, although a potential problem with analogues of insulin that have significant immunogenic differences from endogenous insulin); primary loss of counter-regulatory hormones as in Addison’s disease, hypopituitarism and isolated growth hormone deficiency; and drugs that interfere with endogenous glucose production. In research studies in people with T2DM, where treatment includes but is not exclusively insulin, a history of smoking, lower educational attainment, cognitive function, presence of microvascular disease were all associated with increased severe hypoglycemia, and in one study of hypoglycemia in people on insulin, there was a correlation with socioeconomic status [46,47]. Outside diabetes, hypoglycemia may be a marker of severe illness and elderly people with diabetes and comorbidities are particularly high risk. Drug interactions causing hypoglycemia are more of a problem with people with residual insulin secretion that may be stimulated by agents such as sulfamethoxazole or using oral hypoglycemic agents than with insulin treatment.

Many studies where patients have been randomized to intensive versus conventional or less intensive treatments have shown greater frequency of severe hypoglycemia in the intensive treatment arms. Nevertheless, modern methods of supporting patients in achieving improved diabetes control using structured education can deliver improved glycemic control with reduced risk of severe hypoglycemia (e.g. [12]), and large recent observational studies show greater severe hypoglycemia in those with less efficient control [48]. It may be that unsuccessful attempts to lower HbA1c carry most risk.

**Consequences of hypoglycemia**

An acute hypoglycemic episode can present as a neurologic insult, such as confusion, coma, or seizure but recovery after restoring the blood glucose (or after spontaneous recovery of the blood glucose as active insulin wanes and counterregulation kicks in) is usual. There may be amnesia for the event and musculoskeletal injury can occur as a result of unconsciousness or seizure. Rarely, after nocturnal hypoglycemia, a unilateral paralysis mimicking a stroke may occur (Todd’s paresis) from which recovery occurs over minutes or hours.

Death from hypoglycemia is difficult to quantify and recent assessments of its frequency range from 4–10% of deaths in people with young-onset type 1 disease, presumably the group most at risk. Using the same data, Kerner and Voltze estimated a mortality rate from hypoglycemia of 0.65 per 1000 people, assuming a mean duration of diabetes of 50 years [49]. In insulin-treated type 2, mortality from hypoglycemia is more difficult to quantify. Increased mortality in a randomized
controlled trial of very intensive glycemic control in people with late T2DM and existing cardiovascular disease was associated with increased risk of both severe hypoglycemia and mortality but the link may reflect that hypoglycemia is a marker for higher risk of mortality [46] — a well-evidenced link that may also explain higher mortality in people with T2DM and lower HbA1c whether or not they are on insulin [50].

The common result of a hypoglycemic episode is recovery. In adults, there is little evidence for permanent injury, including cognitive impairment, from hypoglycemic episodes from which an apparently full recovery is made at the time. Cognitive decline has been linked to hypoglycemia but the direction of that link remains debatable, and any relationship is complex [51]. Similarly in T1DM, hyperglycemia rather than hypoglycemia seems to have the stronger association with cognitive impairment [52].

Patients with problematic hypoglycemia may complain of memory deficits. In animals rendered profoundly hypoglycemic hippocampal injury is described and lesions there would affect memory, but it is equally likely on present evidence that the patient is aware of acute effects of hypoglycemia. Memory is not made during hypoglycemia and may not be consolidated during nocturnal episodes. People in the intensive arm of the American Diabetes Control and Complications Trial (DCCT) had better performance on cognitive function tasks years later than those who had been randomized to conventional therapy, despite a much higher incidence of severe hypoglycemic episodes in the latter and this lack of association between severe hypoglycemia and cognitive performance has persisted in later follow-up [53].

In children, severe hypoglycemia, especially if associated with seizure, is associated with measurable impaired function of some cortical domains in later life. Recent studies continue to show mild loss of certain domains of intellectual and cognitive performance in people who have had T1DM, complicated by severe hypoglycemia, in early life, at a time when brain growth is very active (e.g. [54]).

Hypoglycemia may precipitate a cardiac arrhythmia and this is thought to be the agonal event in rare cases of “dead in bed,” a tragic scenario in which a person with T1DM, often young, is found dead in an undisturbed bed. Lengthening of the cQTC interval on the ECG has been seen in both experimentally induced and coincidental clinical hypoglycemia and this is a pro-arythmogenic change. It is associated with the fall in potassium seen with insulin-induced hypoglycemia, and the stress hormone response and can be prevented in an experimental setting by potassium replacement and by beta-blockade [55]. Hypoglycemia is associated with a pro-inflammatory response which includes a hypercoagulatory state. It has been suggested that this may contribute to vascular complications of diabetes, although there is no epidemiologic evidence for this. Vascular complications were reduced during intensive treatment of T1DM in DCCT, despite the higher rate of severe hypoglycemia. In contrast, it may well be that a hypoglycemic event could precipitate an acute myocardial infarction — in the elderly, presentation of severe hypoglycemia is not uncommonly complicated by myocardial infarction or stroke [56].

Socioeconomic impact of hypoglycemia is difficult to quantify and not many studies have restricted themselves to T1DM or insulin-treated diabetes alone. One web-based questionnaire study describes significant socioeconomic impact of nonsevere hypoglycemia associated with a range of diabetes treatments, but also notes very different experiences between different Western cultures [57]. Other studies have failed to find any quality of life impact of nonsevere hypoglycemia. Nevertheless, fear of hypoglycemia remains potent and may interfere with both patient and healthcare willingness to push for tight glycemic control to reduce long-term vascular complications [58].

In contrast, severe hypoglycemia can have significant impact. Injury may occur with unconsciousness or seizure, loss of cognitive function at work may place the patient and others in danger or create embarrassment as the hypoglycemic person loses intellectual control. Severe hypoglycemia and impaired awareness of hypoglycemia can cost people work opportunities and privileges such as a license to drive and impaired quality of life has been reported in a diverse group of people with T1 and T2DM through an online questionnaire [59]. Type 1 patients with impaired awareness report various negative impacts on their lives [60] and one recent study of a small group of people with T1DM and intractable problems with impaired awareness of hypoglycemia and recurrent severe hypoglycemia reported significant stress in their relatives [61]. The cost to the health economy of severe hypoglycemia is not fully explored, as estimates rarely include time off work or costs other than those associated with the immediate event. Even so, a minimum cost has been quoted as in excess of £13 million in the UK [47].

In rare cases, automatism is reported. Automatic behaviors are usually nonpurposeful movements such as lip smacking and associated with epilepsy, but can be more complex, in which the person carries out more sophisticated movements with no awareness or recollection afterwards. In this context, automatism is a legal term, implying loss of voluntary control of behavior. In British law, use of this defense to excuse, for example, a driving accident, will depend on being able to prove that the person had no opportunity to take preventive action — for example eating after administration of insulin, or absence of knowledge that he/she had unawareness of hypoglycemia; in current European law, hypoglycemia unawareness is a barrier to having a license to drive [62].

**Nocturnal hypoglycemia**

In a large glucose-monitoring study of people with T1DM, nocturnal hypoglycemia in people with T1DM complicated about 5–10% of nights [63]. Nocturnal hypoglycemia is, at least 50% of the time, asymptomatic. Studies documenting its frequency continue to report this, despite major changes in insulin regimens and methods of glucose monitoring. As mentioned earlier, in deep sleep, counterregulatory responses to hypoglycemia are significantly attenuated [29], with many episodes being prolonged, and the idea that hypoglycemia at
Hypoglycemia and other complications of insulin therapy

791

night is a common cause for fasting hyperglycemia next day (a reinterpretation of the Somogyi phenomenon) is rarely, if ever, correct. The closest association of a nocturnal hypoglycemia is a fasting (pre-breakfast) glucose of less than 5 mmol L\(^{-1}\) [64]. Although variability of fasting glucose may be associated with risk of nocturnal hypoglycemia, the likely cause of a high pre-breakfast glucose reading is loss of insulin action, perhaps associated with a dawn phenomenon (a rising insulin requirement driven by diurnal variation in hyperglycemic hormones such as growth hormone and cortisol).

Nocturnal hypoglycemia may be associated with reduced mood next morning (although there is no evidence for cognitive impairment), failure to consolidate memory during the day, and reduced responses to hypoglycemia next day [65].

Drivers of nocturnal hypoglycemia include exercise during the day, especially if unusually (for that person) vigorous or prolonged, attributed to depletion of liver and muscle glycogen stores; increased alcohol consumption suppressing gluconeogenesis; and, in insulin-deficient patients, use of conventional rather than analogue insulin [66].

Addressing hypoglycemia in insulin therapy

Hypoglycemia occurs whenever circulating insulin concentrations exceed requirement. The aim of exogenous insulin therapy is to approximate normal physiology but exogenous insulin is not delivered directly to the liver and is not under endogenous homeostatic control.

Assessing a person with diabetes for hypoglycemia experience and risk

Every person using insulin or insulin secretagogues should be made aware of their risk of hypoglycemia and advised about its avoidance, while not engendering inappropriate fear. An assessment of both hypoglycemia experience and risk for severe hypoglycemia should be made at least at annual review and for someone with insulin-deficient (all type 1 and late type 2) diabetes at every therapeutic encounter. Patients may define their events very differently from the healthcare professional, describing as severe an unpleasantly symptomatic episode, which poses a lower health risk than do asymptomatic episodes. People should be specifically asked about their experience of hypoglycemic episodes for which they required help (severe hypoglycemia), documenting the time of day, possible precipitants and outcomes of each one since the last visit; and likewise documenting frequency and timing of episodes that they recognized and treated themselves (Table 53.2). Nocturnal episodes and night sweats that might represent otherwise unrecognized hypoglycemia should be documented, with other causes of night sweating (e.g., chronic infection, disseminated malignancy and lymphoma, autonomic neuropathy, thyrotoxicosis, an overly warm environment) considered. Asymptomatic hypoglycemia is best described by patients’ relatives but can be detected by asking about episodes where other people tell the patient he/she is hypoglycemic before they have recognized it for themselves or by coincidental home blood glucose measurement. Again, if present, frequency and high-risk time of day should be documented. Qualitative assessment of hypoglycemia awareness should also be made, using one of the questionnaires validated in terms of usefully predicting severe hypoglycemia. The Gold score [67] simply asks the patient to rank their hypoglycemia awareness from always aware (1) to never aware (7) and defines impaired awareness as a score of 4 or more. The Clarke score [68] is a more complex questionnaire which includes data on frequency of severe hypoglycemia, frequency of episodes of exposure to hypoglycemia with or without symptoms, and the glucose concentrations perceived as hypoglycemic. Another simple question which has validity in terms of being associated with severe hypoglycemia risk is to ask whether the patient is aware of hypoglycemia at glucose concentrations of 3 mmol L\(^{-1}\) or more, under 3 mmol L\(^{-1}\), or not at all, with the latter two being classed as impaired awareness [12]. A more complex but well-validated assessment of hypoglycemia burden includes documenting all episodes over a month of observation [69]. Patients with complete hypoglycemia unawareness should be warned against driving and those with incomplete impaired awareness cautioned to use a higher glucose target as indicating safety to drive than those with good awareness, for example 7 mmol L\(^{-1}\) (126 mg dL\(^{-1}\)), as blood glucose falls during driving.

Having assessed hypoglycemia risk, the person who is experiencing problems (more than two symptomatic (mild) episodes per week; reduced awareness; recurrent severe episodes) should be investigated. Initial questions depend upon presentation and minor hypoglycemia, or hypoglycemia with an obvious precipitant, may be dealt with by insulin dose adjustment and/or education in insulin dose flexing around, for example, exercise or alcohol intake. A severe hypoglycemic event occurring during a skiing holiday needs a different approach from severe hypoglycemia occurring randomly. Recurring precipitants of hypoglycemic events should be sought by history, assisted by inspection of patient records of glucose monitoring. If nothing in the pattern suggests a remedial regimen error or lifestyle cause, the person should be checked for health issues placing them at particularly high risk (Table 53.1). These are rare but need to be ruled out before focusing exclusively on improving the insulin replacement regimen. Replacement of deficient growth hormone secretion as a treatment for problematic hypoglycemia in an adult with diabetes carries a risk of exacerbating diabetic retinopathy and should only be done cautiously and with regular monitoring of the retinae.

Management of hypoglycemia

Acute episodes

The correct treatment of a hypoglycemia episode is immediate ingestion of a carbohydrate that will rapidly elevate the circulating glucose concentrations. People with insulin-treated
### Table 53.2 Assessing individual hypoglycemia exposure

<table>
<thead>
<tr>
<th>Documentation of experience</th>
<th>Number since last visit</th>
<th>Time of day of each</th>
<th>Possible precipitants</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Severe episodes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Requiring assistance</td>
<td></td>
<td>Premeal</td>
<td>Insulin error</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post meal</td>
<td>Missed meal</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nocturnal</td>
<td>After increased</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>exercise</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>After increased</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>alcohol</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Other comorbidity</td>
</tr>
<tr>
<td><strong>Mild (symptomatic)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-treated</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Asymptomatic episodes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Routine SMBG readings</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;3 mmol L(^{-1}) with no symptoms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Others noticing hypoglycemia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Nocturnal episodes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number per month</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Possible precipitants</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Other night sweats</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number per month</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rank ability to feel hypoglycemia by symptoms from 1 (all events) to 7 (never) (^a)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Always aware</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>Never aware</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)Gold score, [67].

diabetes should always carry something suitable. Mild episodes can be treated by consumption of 15–20 gm “rapidly available” carbohydrate in the form of 100 mL of non-diet conventional Lucozade, 200 mL fresh fruit juice or, more conveniently for portability, glucose tablets. In the UK, five Bassett’s jelly babies currently constitute just over 20 gm — six may be needed with smaller versions. If a meal is not due within the next half hour, a further 20 gm of more slowly absorbed carbohydrate — two semisweet biscuits for example—should also be taken to maintain the glucose concentrations. If a meal is due, the hypoglycemia should be corrected as above, with the usual dose of insulin injected premeal, with a dose reduction if the cause of the original hypoglycemia is still present — for example after exercise. Delaying the meal injection will only drive immediate postprandial hyperglycemia — the meal carbohydrate content still needs to be managed — and increase the risk of further hypoglycemia later on as the insulin effect continues. Missing the dose altogether predictably results in hyperglycemia. Patients experiencing hypoglycemia sometimes struggle to restrict their carbohydrate intake to the recommended amounts and post-hypoglycemic hyperglycemia may ensue. This should not be corrected with further insulin injection.

Glucose gels are available for buccal administration but if a person is confused, and there are concerns that he/she may not be able to swallow safely, parenteral therapy is recommended. In the home, this can be 1 mg glucagon injected intramuscularly, which will restore circulating glucose concentrations within 10 minutes. Glucagon will not work if hepatic glycogen
stores are depleted, for example within 24 hours of an earlier glucagon-treated hypoglycemia; in liver failure; in cases of severe anorexia or malnutrition, including that associated with long-term severe illness; or if endogenous glucose production is otherwise suppressed, for example in alcohol intoxication. For these reasons, in hospital, intravenous glucose is generally preferred. There is no indication for highly concentrated glucose solutions except in people with acutely compromised circulation and a central line in place. Larger volumes of less concentrated glucose (e.g. 150 mL 10%) are much safer and equally effective. Extravasated 50% glucose can cause local tissue necrosis. Families should not be given responsibility for intravenous therapy. A person with diabetes requiring intravenous glucose needs to be in a properly regulated hospital environment.

After treatment, plasma glucose should be checked within 30 min to ensure recovery. Patients, and healthcare professionals, should be made aware that the risk for a second hypoglycemic event is greatly increased within 24 hours of the first. This is presumably in part because of continuation of the stimulus for the first hypoglycemia, in part because of depletion of glucose stores in liver in the response to the initial hypoglycemia, and in part because of the hypoglycemia-induced impairment of counterregulation described earlier.

**Prevention of future events and restoration of impaired hypoglycemia awareness**

Having established that a person has problems with hypoglycemia, and that no comorbidity needs treating, the patient's self-management regimen should be checked. Structured education in flexible insulin therapy using multiple daily injections has repeatedly been shown to reduce severe hypoglycemia and even restore impaired awareness, presumably as more physiologic insulin replacement, together with a strategy for adjusting basal as well as meal replacement doses when increasing exercise or alcohol intake, reduces the occurrence of any hypoglycemia [12]. There are many such programs, all teaching people how to assess carbohydrate content of desired food and determine the amount of insulin they require for each unit of carbohydrate based on observation of blood glucose monitoring. They also teach how to adjust basal insulin requirement, to control endogenous glucose production, based on blood tests and lifestyle issues such as activity levels. The UK's Dose Adjustment For Normal Eating (DAFNE) program reports that of the 40% of people coming for training with impaired hypoglycemia awareness, 43% have restored awareness 1 year later. A small number, however, develop impaired awareness apparently *de novo* during the year, leaving 30% with impaired awareness (cf. 40% at entry) at 1 year [12]. This improvement in awareness status is accompanied by a major reduction in severe hypoglycemia rate, associated with significant improvement of glycated hemoglobin, confirming the original observation that such structured education could durably reduce HbA1c and severe hypoglycemia rate at the same time.

Common behaviors that increase hypoglycemia risk that can be addressed in educational programs include failing adequately to reduce insulin active during and/or in the night following exercise or increased alcohol intake; post-meal injection of meal-related insulin; overcorrection of intermittent high blood glucose, for example within 2 hours of eating or on a high but falling glucose following an earlier correction. Fear of modest hyperglycemia may underpin some of these activities.

Some educational packages have been developed specifically targeting avoidance of extremes of blood glucose and concentrate on teaching people techniques to predict their blood glucose (based on a knowledge of their insulin’s pharmacodynamics and its interactions with food, exercise, etc.) and better ways of recognizing symptoms of impending hypoglycemia. These have been documented also to reduce hypoglycemia burden [70].

For people who cannot or will not attend such programs, using their principles to reduce hypoglycemia risk is useful, although this has not been formally demonstrated. Ensuring adequate and stable background insulin replacement, for example using twice daily administration of basal insulin, allows flexibility around changes in lifestyle and reduces dependence on meal insulin to provide also insulinization between meals. Taking the evening basal insulin at bedtime, rather than before the evening meal, delays the peak action of the insulin till the time of maximum insulin sensitivity is passed and use of “peekless” basal analogues further diminishes the risk of nocturnal hypoglycemia in patients needing full insulin replacement, as well as providing more stable control of fasting glucose [71]. Use of fast-acting insulin analogues injected 10–15 minutes pre-meal also reduces late hypoglycemic drive, compared to the pre-meal injection of soluble (regular) animal or human insulins, which have a relatively longer duration of action, when between-meal snacking may be required to achieve optimal control. The impact of rapid-acting insulin analogues on nocturnal hypoglycemia suggests that nocturnal hypoglycemia has been driven in part by the tail of the pre-evening meal soluble insulin, which can be offset by bedtime snacking. Flexible insulin therapy training favors adjustment of meal doses according to the amount of carbohydrate the person proposes to eat, and many programs, with careful attention to background as well as meal-related insulin dosing, run without prescribing routine between meal snacking, perhaps because of reduced dependence on the meal insulin to provide basal cover between meals. Routine snacking may become useful in people struggling to achieve good glycemic control because of hypoglycemia before the next meal, assuming the basal contribution has been assessed as appropriate, for example in pregnancy, when there is a need for strict control of pre- and postprandial glucose concentrations.

For people whose problems with hypoglycemia persist, replacing multiple daily injection therapy with continuous subcutaneous insulin infusion (CSII), has been shown to achieve a fourfold reduction in severe hypoglycemia risk, in a meta-analysis of studies done mostly in people with clinical indications for pump therapy [72]. It is not known why CSII
has such benefit on severe hypoglycemia rates but its ability to accommodate a dawn phenomenon with reduced nocturnal hypoglycemia may be important, as may the incorporation of devices to assist in bolus calculations that take into account earlier insulin administration. CSII is also associated with improving HbA1c, which should finally break the idea that intensifying insulin therapy and the achievement of optimal glucose control for prevention of structural complication are inextricably linked. “Loosening” diabetes control to deal with hypoglycemia risk simply exchanges one problem for another and the correct management of the person who is hypoglycemia-prone is a reinforcement of the importance of adjusting therapy to stay below the lower limit of the target range.

Early studies of real-time online continuous glucose monitoring excluded people at high risk for hypoglycemia, so event rates were too small to give definitive answers for its ability to prevent hypoglycemia but they may provide benefit [73] especially for those at high risk [74]. A high degree of compliance with sensors may be required for them to provide benefit. A system that allows detection of hypoglycemia by a sensor to drive an automatic suspension of insulin delivery does reduce duration of nocturnal hypoglycemia in those at highest risk and recently has been shown dramatically to reduce severe hypoglycemia in a high-risk group [75]. Closed-loop insulin delivery systems may have more to offer but remain in development at the time of writing.

Transplantation as a treatment for hypoglycemia

For people who cannot stop having severe hypoglycemia despite all of the aforementioned treatments, the option of islet cell replacement by islet cell transplantation is available in some countries. Replacement of functional β cells seems to provide robust protection against severe hypoglycemia in people at very high risk, even if insulin independence is not achieved. The international transplant registry reported that 67% of islet recipients have detectable C-peptide and no hypoglycemia 5 years post transplant, while only 13% of recipients were still insulin independent [76]. Research is developing the techniques involved, from the isolation and manipulation of the islets, to the immunosuppression and anticoagulation used to protect them post transplant and it is expected that these results will improve over time. Whole organ pancreas transplant is another option, discussion of which is beyond the scope of this chapter. Meanwhile, while this treatment is for only a few people worldwide at present, lessons can be learned about how it defends against hypoglycemia, which may be more widely relevant. There are no randomized controlled trials of islet transplantation versus other therapies to reduce hypoglycemia but the people being accepted into islet programs almost all have a long history of hypoglycemia unawareness and high burden of hypoglycemia experience. Possible factors in the success of islet replacement include the replacement of some endogenous insulin which can be suppressed during hypoglycemia and/or a role for C-peptide to have some protective action.

Barriers to avoidance of hypoglycemia and restoration of impaired hypoglycemia awareness

There is evidence to suggest that people with T1DM and impaired hypoglycemia awareness may be less compliant with suggested changes in their self-management regimens than those with preserved awareness. This may relate to abnormal pre-frontal responses to hypoglycemia described previously [45]. In interviews with people experiencing impaired awareness, Rogers found a minority expressed high concern about their condition, with more expressing less concern and less motivation to regain awareness [77], and patients with impaired awareness are less compliant with therapeutic change [78]. It is important to note that a small number of these people did not seem to have a significant clinical problem with hypoglycemia, suggesting that their low level of concern was warranted. The remainder reported less concern despite describing very serious outcomes of experienced hypoglycemia in the past. These people had different cognitions around their hypoglycemia which prevented them engaging in preventive strategies. Among other important issues was the belief that their hypoglycemia could not be prevented, an attitude that may be shared by healthcare professionals. In order to engage people with relatively low concern in strategies to diminish their hypoglycemia, it may be necessary to use techniques specifically directed at motivation and cognitive change [61], although the success of these approaches requires further validation.

Nonhypoglycemic complications of insulin therapy

Weight gain

Weight gain with insulin therapy is a second barrier to achieving good glycemic control and this can be an almost insuperable barrier in some people. People admit to omitting insulin to control weight especially when young. Weight gain may be a particular problem when attempting to improve diabetes control that has been very poor. In that setting, sometimes at diagnosis after weight loss, and sometimes after periods of poor control later in the course of the disease, weight gain is multifactorial [79]. If re-insulinizing after a period of significant insulin deprivation, salt and water retention may result in clinical edema with weight gain. This needs active management and the patient should be encouraged that the effect is transient.

More difficult is the weight gain associated with the retention of calories previously lost in glycosuria. It is claimed that 500 kcal per day can be lost during poor diabetic control through this route. Patients striving to achieve improved glycemic
control need to be aware of this and may be helped by using weight-controlling diets and exercise to offset it. The role of anti-obesity agents in this context has not been explored.

Other contributors to weight gain with insulin may include a need to eat to prevent or treat hypoglycemia—this should be addressed by revision of the insulin regimen and how the patient is using it. Although in animal studies, insulin in the brain acts to suppress appetite, patients commonly describe insulin as driving them to eat more. It is possible that some insulins have less effect on appetite than others and insulin detemir as a basal insulin in T1DM is associated with reduced food intake and weight gain compared to NPH insulin. This field also requires more investigation.

Weight gain may be associated with other poor health outcomes, such as increased cardiovascular risk. Indeed, as insulin resistance is associated with peripheral hyperinsulinemia and exogenous insulin therapy may also be (see earlier), it has sometimes been suggested that insulin therapy itself may increase cardiovascular risk. An early study of patients in the DCCT showed increased cardiovascular risk in those subjects in the intensive arm who gained most weight. In a long-term follow-up of DCCT participants, significantly less hypertension was found in patients who had received intensive insulin therapy during the DCCT, although it is noted that weight gain in these patients did reduce the effect [80].

Injection site issues

Insulin allergy

Insulin allergy is fortunately rare and commonly directed towards the noninsulin components of insulin preparations. It may present as a local response at the site of injection, with a classical delayed hypersensitivity response, or more rarely manifest as a systemic reaction and occasionally anaphylaxis. Patients can be tried on different insulin formulations but skin testing and blood tests for anti-insulin immunoglobulins should be considered to make a formal diagnosis. Treatment strategies have included use of antihistamine therapies and steroids (both locally in the injection and systemically) and insulin pump therapy, as the continuous exposure to low doses of the allergen appear to induce tolerance. In resistant serious cases, use of immunosuppressive agents to deplete B lymphocytes has been reported [81].

Lipohypertrophy and lipoatrophy

Lipohypertrophy is a local overgrowth of subcutaneous adipose tissue in response to the lipotrophic actions of insulin. It occurs when a patient consistently injects into a site, and is probably very common. The natural tendency is to inject into a site that is routinely available and patients may concentrate on one site as it becomes less sensitive to the sensation of injection. The swelling may be subtle but is poorly vascularized and this may interfere with the efficiency of the injected insulin and/or its pharmacodynamics. Where carried out to extreme, unsightly swellings can occur and insulin absorption may be significantly affected. Treatment is to avoid further injection into the site and to rotate injection sites thereafter so as not to recreate the problem elsewhere.

Lipoatrophy describes pitting of the skin as the subcutaneous tissue is reduced. It is thought to result from immune complex formation and its occurrence reduced with the introduction of highly purified animal and human insulins, which are less immunogenic than older forms of insulin. It does not seem to have made a reappearance with insulin analogues. Treatment is by “flooding” the affected area with excess antigen, which can be achieved by asking the patient deliberately to inject a purer insulin into the affected site.

Other skin lesions may manifest as a result of poor injection technique. Infection is rare but bleeding, pain, and bruising may require review of technique and sometimes conversion to a different needle length. Usually this will be to a shorter needle, but if insulin leakage is a problem, an 8 mm needle may be better than a shorter one. Intradermal injection can cause small skin lesions, although this may be less of a problem with modern needles.

Fear of injection is multifactorial and in patients with T2DM delaying converting to insulin, reassurance about the natural course of T2DM, and the modern trend to early insulin introduction may help. Some natural anxiety around injections may be resolved after the first self-administration proves not to be painful but cultural issues may complicate patients’ acceptance of the need to inject. A true needle phobia should be suspected if the patient describes taking a long time over each injection and injection avoidance. Where suspected, a true needle phobia should be referred for psychiatric assessment and psychologic therapies, although conversion to CSII, as part of the therapy, may help.

Cancer

Certain types of cancer have been associated with insulin resistance and it has been suggested that exogenous insulin therapy may carry similar risk though peripheral hyperinsulinemia. There is no hard evidence to support this claim which is more fully discussed in Chapter 21 [82].

References


CHAPTER 54
Diabetic ketoacidosis and hyperosmolar state

Abbas E. Kitabchi\textsuperscript{1}, Guillermo E. Umpierrez\textsuperscript{2} and Mary Beth Murphy\textsuperscript{1}
\textsuperscript{1}The University of Tennessee Health Science Center, Memphis, TN, USA
\textsuperscript{2}Emory University School of Medicine, Atlanta, GA, USA

Key points

- Admission for DKA and HHS may be increasing due to economic and psychosocial stressors.
- Mortality rate is higher for HHS than DKA and is characterized by marked stupor without acidosis and extreme elevations of blood glucose with severe dehydration.
- There are three distinct components of DKA: hyperglycemia, acidemia, and ketonemia.
- DKA is classified as mild, moderate, or severe by the American Diabetes Association with options for treatment locations depending on severity.
- Most patients with DKA have autoimmune type 1 diabetes but patients with type 2 diabetes are also at risk for DKA during catabolic stress.
- Hyperglycemic crisis is a pro-inflammatory state with elevations of cardiovascular risk factors, oxidative stress, and pro-inflammatory cytokines which decline after resolution of DKA.
- More than half of obese African Americans with newly diagnosed diabetes presenting with DKA have no apparent precipitating cause. These patients may be labeled idiopathic T1DM, atypical diabetes mellitus, and ketosis-prone T2DM. Interestingly, 40% of these patients remain non-insulin-dependent 10 years after diabetes onset and display a nonautoimmune type of diabetes.
- Common precipitating causes for DKA include: infections, intercurrent illness, psychological stress, and noncompliance with therapy.
- Treatment of DKA includes improvement of organ perfusion by increasing circulatory volume, gradual reduction of osmolality and glucose, clearance of serum and urine ketones, and normalizing electrolyte levels.
- After recovery of DKA, intravenous infusion of insulin should continue for at least 1 hour while transitioning from intravenous to subcutaneous insulin therapy to prevent regression to DKA.
- Altered sensorium in DKA is independently associated with acidosis.

Introduction

Diabetic ketoacidosis (DKA) and hyperglycemic hyperosmolar state (HHS) are the most serious acute metabolic emergencies in type 1 (T1DM) and type 2 diabetes (T2DM). Metabolic derangements result from the combination of absolute or relative insulin deficiency and an increase in counterregulatory hormones (glucagon, catecholamines, cortisol, and growth hormone). Successful treatment of DKA requires frequent monitoring of circulatory volume and tissue perfusion, correction of hypovolemia and hyperglycemia, replacement of electrolyte losses, and a careful search for the precipitating cause. In spite of advances in our understanding of the underlying mechanisms that lead to the development of hyperglycemic crises, DKA continues to be an important cause of morbidity and mortality among patients with diabetes.

The number of hospital admissions for DKA in the US is over 130,000 per year and has shown an upward trend, with a 30% increase in the annual number of cases between 1995 and 2005 \cite{1}. Treatment of DKA utilizes a large number of resources, accounting for an estimated total cost of $2.4 billion annually \cite{1}. In adult patients, recent controlled studies have reported a mortality rate less than 2%, with higher mortality observed in elderly subjects and in patients with concomitant life-threatening illnesses.

This chapter reviews recent advances on the epidemiology, diagnosis, pathogenesis, and current recommendations based on the most recent American Diabetes Association (ADA) guidelines for the management of adult patients with DKA \cite{2}.

Diabetic ketoacidosis

Definition and classification

Diabetic ketoacidosis consists of three distinct components: hyperglycemia, acidemia, and ketonemia. Each of these components may be caused by a variety of conditions. According to the recent ADA guidelines, DKA is classified as mild, moderate, or severe based primarily on the blood pH, bicarbonate, ketones, and alteration of sensorium (Table 54.1) \cite{2}. Table 54.1
Table 54.1 Diagnostic criteria and typical total body deficits in DKA and HHS

<table>
<thead>
<tr>
<th>Diagnostic criteria and classification</th>
<th>Mild</th>
<th>Moderate</th>
<th>Severe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plasma glucose, mg dL⁻¹</td>
<td>&gt;250</td>
<td>&gt;250</td>
<td>&gt;250</td>
</tr>
<tr>
<td>Arterial pH</td>
<td>7.25–7.30</td>
<td>7.00&lt;7.24</td>
<td>&lt;7.00</td>
</tr>
<tr>
<td>Serum bicarbonate, mEq L⁻¹</td>
<td>15–18</td>
<td>10&lt;15</td>
<td>&lt;10</td>
</tr>
<tr>
<td>Urine ketone</td>
<td>Positive</td>
<td>Positive</td>
<td>Positive</td>
</tr>
<tr>
<td>Serum ketone</td>
<td>Positive</td>
<td>Positive</td>
<td>Positive</td>
</tr>
<tr>
<td>Effective serum osmolality</td>
<td>Variable</td>
<td>Variable</td>
<td>Variable</td>
</tr>
<tr>
<td>Anion gap</td>
<td>&gt;10</td>
<td>&gt;12</td>
<td>&gt;12</td>
</tr>
<tr>
<td>Mental status</td>
<td>Alert</td>
<td>Alert/drowsy</td>
<td>Stupor/coma</td>
</tr>
</tbody>
</table>

Typical deficits

<table>
<thead>
<tr>
<th></th>
<th>DKA</th>
<th>HHS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total water, L</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>Water, mL kg⁻¹</td>
<td>100</td>
<td>100–200</td>
</tr>
<tr>
<td>Na⁺, mEq kg⁻¹</td>
<td>7–10</td>
<td>5–13</td>
</tr>
<tr>
<td>Cl⁻, mEq kg⁻¹</td>
<td>3–5</td>
<td>5–15</td>
</tr>
<tr>
<td>K⁺, mEq kg⁻¹</td>
<td>3–5</td>
<td>4–6</td>
</tr>
<tr>
<td>PO₄, mmol kg⁻¹</td>
<td>5–7</td>
<td>3–7</td>
</tr>
<tr>
<td>Mg²⁺, mEq kg⁻¹</td>
<td>1–2</td>
<td>1–2</td>
</tr>
<tr>
<td>CA++, mEq kg⁻¹</td>
<td>1–2</td>
<td>1–2</td>
</tr>
</tbody>
</table>

*Euglycemic DKA has been reported [22,23].
†Nitroprusside reaction method.
‡Effective serum osmolality calculation: 2[measured Na⁺ (mEq L⁻¹)] + glucose (mg dL⁻¹)/18[mOsm kg⁻¹].
§Anion gap calculation: (Na⁺) - (Cl⁻ - HCO₃⁻) (mEq L⁻¹) [normal = 12 ± 2].
¶Per kilogram of body weight.

also provides the criteria for the diagnosis of hyperglycemic hyperosmolar state (HHS).

**Epidemiology**

Observational studies have reported that DKA accounts for 4–9% of hospital discharges among patients admitted with a primary diagnosis of diabetes [3]. A higher figure was reported by the EURODIAB study where 8.6% of 3250 subjects with T1DM throughout Europe had been admitted with DKA in the previous 12 months [4]. Recent epidemiologic studies indicate that hospitalizations for DKA during the past two decades are increasing, with the majority of cases occurring as recurrent cases in the same subjects [1,3].

Most patients with DKA have autoimmune T1DM; however, patients with T2DM are also at risk during the catabolic stress of acute illness such as trauma, surgery, or infection. In contrast to popular belief, DKA is more common in adults than in children. In community-based studies more than 40% of patients with DKA are older than 40 years and more than 20% are older than 55 years [5]. More than half of obese African Americans with newly diagnosed diabetes presenting with DKA have no apparent precipitating cause [6]. At presentation, obese T2 patients with DKA have markedly impaired insulin secretion and insulin action [6]. Intensified diabetic management results in significant improvement in β-cell function and insulin sensitivity sufficient to allow discontinuation of insulin therapy within a few months of follow-up. This clinical presentation has been reported primarily in African Americans, but also in other minority ethnic groups, and has been labeled idiopathic T1DM, atypical diabetes mellitus, type 1.5 diabetes, and more recently as ketosis-prone T2DM [6,7]. Despite their presentation in DKA, these patients have T2DM because of the presence of obesity, a strong family history of diabetes, measurable insulin secretion, and a low prevalence of autoimmunity markers of β-cell destruction [7,8]. Furthermore, 40% remain non-insulin-dependent 10 years after diabetes onset and display a nonautoimmune type of diabetes [9].

In adult subjects with DKA, the overall mortality is less than 1%; however, a mortality rate higher than 5% has been reported in the elderly and in patients with concomitant life-threatening illnesses [2,5]. Death in these conditions is rarely due to the metabolic complications of hyperglycemia or ketoacidosis, but relates to the underlying precipitating illness. The prognosis of both conditions is substantially worsened at the extremes of age, in the presence of coma, hypotension, and severe comorbidities.

**Precipitating factors**

In patients with known diabetes, precipitating factors for DKA include infections, intercurrent illnesses, psychological stress, and noncompliance with therapy. Worldwide, infection remains the most common underlying cause, occurring in 30–50% of cases [2]. Urinary tract infection and pneumonia account for the majority of infections [10]. Other acute conditions
that may precipitate DKA include cerebrovascular accident, alcohol abuse, pancreatitis, pulmonary embolism, myocardial infarction, surgery, and trauma. Drugs that affect carbohydrate metabolism such as corticosteroids, sympathomimetic agents, and pentamidine may also precipitate DKA. Recently a number of case reports indicate that conventional as well as atypical antipsychotic drugs produce diabetes [11].

In adult patients with T1DM, poor adherence to insulin therapy is reported as the major precipitating cause of DKA in inner city populations [12]. A recent study determined that clinical, socioeconomic, and psychological factors were associated with recurrence of DKA in inner city minority patients [13]. Discontinuation of insulin therapy accounted for over two thirds of all DKA admissions. Several behavioral, socioeconomic, and psychosocial factors contributed to poor treatment adherence. Among patients with poor compliance with insulin therapy, one third of patients “just stopped” insulin with no clear explanation, one third reported financial troubles, and most of the rest reported being away from supply or did not know how to handle insulin on sick days. Other studies have reported that psychological risk factors including eating disorders in up to 20% of recurrent episodes of ketoacidosis in young women [14]. Recently, it was estimated that up to one third of young women with T1DM have eating disturbances [15].

The most common precipitating or predisposing factors for the development of HHS is an infectious process or an intercurrent medical or surgical illness [10,16]. Table 54.2 lists the common predisposing or precipitating factors for the development of HHS.

### Table 54.2 Predisposing or precipitating factors for HHS

<table>
<thead>
<tr>
<th>Category</th>
<th>Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute illness</td>
<td>Acute infection (32–60%)</td>
</tr>
<tr>
<td></td>
<td>Pneumonia</td>
</tr>
<tr>
<td></td>
<td>Urinary tract infection</td>
</tr>
<tr>
<td></td>
<td>Sepsis</td>
</tr>
<tr>
<td></td>
<td>Cerebral vascular accident</td>
</tr>
<tr>
<td></td>
<td>Myocardial infarction</td>
</tr>
<tr>
<td></td>
<td>Acute pancreatitis</td>
</tr>
<tr>
<td></td>
<td>Acute pulmonary edema</td>
</tr>
<tr>
<td></td>
<td>Intestinal obstruction</td>
</tr>
<tr>
<td></td>
<td>Dialysis, peritoneal</td>
</tr>
<tr>
<td></td>
<td>Mesenteric thrombosis</td>
</tr>
<tr>
<td></td>
<td>Renal failure</td>
</tr>
<tr>
<td></td>
<td>Heat stroke</td>
</tr>
<tr>
<td></td>
<td>Hypothermia</td>
</tr>
<tr>
<td></td>
<td>Subdural hematoma</td>
</tr>
<tr>
<td></td>
<td>Severe burns</td>
</tr>
<tr>
<td>Endocrine</td>
<td>Acromegaly</td>
</tr>
<tr>
<td></td>
<td>Thyrotoxicosis</td>
</tr>
<tr>
<td></td>
<td>Cushing syndrome</td>
</tr>
<tr>
<td>Drugs/therapy</td>
<td>β-Adrenergic blockers</td>
</tr>
<tr>
<td></td>
<td>Calcium channel blockers</td>
</tr>
<tr>
<td></td>
<td>Chlorpromazine</td>
</tr>
<tr>
<td></td>
<td>Cimetidine</td>
</tr>
<tr>
<td></td>
<td>Diazoxide</td>
</tr>
<tr>
<td></td>
<td>Diuretics</td>
</tr>
<tr>
<td></td>
<td>Encainide</td>
</tr>
<tr>
<td></td>
<td>Ethacrylic acid</td>
</tr>
<tr>
<td></td>
<td>Immunosuppressive agents</td>
</tr>
<tr>
<td></td>
<td>L-Asparaginase</td>
</tr>
<tr>
<td></td>
<td>Loxapine</td>
</tr>
<tr>
<td></td>
<td>Phenytoin Propranolol</td>
</tr>
<tr>
<td></td>
<td>Steroids</td>
</tr>
<tr>
<td></td>
<td>Total parenteral nutrition</td>
</tr>
<tr>
<td></td>
<td>Previously undiagnosed diabetes</td>
</tr>
</tbody>
</table>


Pathogenesis

DKA is a state of severe metabolic decompensation characterized by hyperglycemia, metabolic acidosis, and increased total ketone bodies or ketoacids. Ketoacidosis results from insulin deficiency and excess counterregulatory hormones including glucagon, catecholamines, cortisol, and growth hormone. The insulin deficiency of DKA can be absolute in T1DM or relative as in T2DM in the presence of stress or intercurrent illness that causes sudden worsening of insulin resistance and impairment of insulin secretion [16]. In DKA, reduced effective insulin concentrations and increased concentrations of counterregulatory hormones (catecholamines, cortisol, glucagon, and growth hormone) lead to an abnormal metabolism of carbohydrate, protein, and fat as well as disturbances of fluid and electrolyte balances (Figure 54.1). The temporal relationship of the secretion of counterregulatory hormones which leads to the genesis of metabolic decompensation in DKA is the subject of debate. Data on the role of these hormones are based on the results of studies on acute insulin withdrawal in patients with T1DM. Insulin withdrawal in patients with T1DM leads to activation of protein-kinase-induced stimulation of hormone-sensitive lipase in adipose tissue [16]. The increased activity of tissue lipase causes breakdown of triglycerides to glycerol and free fatty acids (FFA). The increased availability of glycerol is utilized as a substrate for gluconeogenesis in the liver and kidney, whereas the massive release of FFA serves as the main precursor of ketoacids in the liver [17]. Increased production of FFA plays a pivotal role in ketogenesis in DKA [18], by suppressing the glycolytic pathway through the inhibition of the rate-limiting enzymes of glycolysis (hexokinase, phosphofructokinase, and pyruvate kinase) as well as pyruvate dehydrogenase; increasing production of fatty acyl-CoA, which (a) inhibits the rate-limiting enzyme citrate synthetase, the first step in the TCA (tricarboxylic acid) Krebs cycle, and (b) stimulates hyperlipidemia through conversion to triacylglycerol; and (c) increases β-oxidation of FFA leading to increased ketogenesis (Figure 54.1).

Carbohydrate metabolism

As a result of increased counterregulatory hormones, particularly glucagon and catecholamines, coupled with a reduction of
insulin, five major pathways of carbohydrate metabolism are affected in DKA [16]:

1. glycogen synthesis is decreased secondary to reduction in glycogen synthase as a result of insulin deficiency;
2. glycogenolysis is accelerated as a result of stimulation of glycogen phosphorylase through catecholamines’ β-receptor stimulation and cyclic AMP; glycogen is then broken down to glucose-1-phosphate;
3. inhibition of glycolysis, as a result of FFA-induced inhibition of rate-limiting glycolytic enzymes (see earlier);
4. decreased glucose utilization by peripheral tissues as a result of insulin deficiency;
5. stimulation of gluconeogenesis by activation of four rate-limiting enzymes: (i) pyruvate carboxylase, (ii) phosphoenolpyruvate carboxykinase, (iii) fructose-1,6-biphosphatase, and (iv) glucose-6-phosphatase.

Additionally, increased levels of alanine (in the liver) and glutamine (from the kidney), as a result of increased proteolysis [16,19]; lactate, as a result of decreased glucose oxidation; and glycero from increased lipolysis serve as substrates for gluconeogenesis. The hepatic glucose production in DKA may be as much as 500 g·d⁻¹. Although gluconeogenesis occurs mainly in the liver, the kidneys also contribute to this process in patients with DKA [16]. An increased glucagon/insulin ratio also (a) inhibits the level of fructose-2,6-biphosphatase leading to a stimulation of fructose-1,6-biphosphatase, (b) stimulates phosphoenolpyruvate carboxykinase, and (c) stimulates glucose-6-phosphatase. The combination of the above pathways leads to hyperglycemia in DKA (Figure 54.1).

**Fat and ketone body metabolism**

With decreased availability of insulin, the major antilipolytic hormone, and increased counterregulatory hormones such as catecholamines, glucagon, cortisol, and growth hormone, there is an accelerated breakdown of triglycerides to glycerol and FFA through activation of hormone-sensitive lipase [16]. Glycerol, which serves as a carbon skeleton for gluconeogenesis and accelerated production of FFA, leads to its further β-oxidation to ketone bodies. This process is predominantly stimulated by glucagon [17,18]. FFA are also converted to triacylglycerol, which may lead to increased very low density lipoprotein production and hyperlipidemia [18]. In DKA, the combination of increased glucagon and low insulin levels decreases liver malonyl-CoA because of a reduction in hepatic carnitine. Reduced malonyl-CoA will permit activation of the rate-limiting enzyme of ketogenesis, carnitine palmitoyltransferase 1 (CPT-1), which allows transesterification of fatty acyl carnitine and oxidation of FFA to ketone bodies [16,18]. CPT-1 is necessary to move FFA into the mitochondria. In addition, metabolism and clearance of these bodies are decreased in DKA. These biochemical pathways are depicted in Figure 54.1.

**Protein and amino acid metabolism**

Negative nitrogen balance is the hallmark of DKA [16]. Several studies have shown that severe insulin deficiency results in a loss of 9–12 g of nitrogen per day, with slow recovery after insulin therapy. Furthermore, it was shown that during insulin withdrawal, the levels of gluconeogenic amino acids (glutamine, alanine, threonine, serine, glutamate, and glycine)
were decreased but levels of ketogenic amino acids (such as leucine, valine, and isoleucine) were increased during DKA [16]. It has been suggested that both increased proteolysis and decreased protein synthesis are responsible for these observations, but the detailed mechanism for these steps is not clearly elucidated. However, increased levels of glucagon and cortisol appear to play major synergistic roles.

**Water and electrolyte metabolism**

Table 54.1 illustrates electrolyte and water deficits in DKA as well as in HHS [10]. The severe derangement of water and electrolytes in DKA and HHS is the result of insulin deficiency, hyperglycemia, and hyperketonemia. Hyperglycemia leads to glucosuria and osmotic diuresis when the renal threshold of glucose (~200 mg dL⁻¹) is exceeded. The osmotic effects of glucosuria result in impairment of NaCl and water reabsorption in the proximal tubule and loop of Henle, resulting in loss of water, sodium, potassium, as well as other electrolytes [19]. In addition, insulin deficiency per se may also contribute to renal losses of water, sodium, potassium, and ammonium salts because insulin stimulates salt and water reabsorption in the proximal and distal nephron. During the development of DKA, intracellular dehydration also occurs as hyperglycemia and water loss lead to increased plasma tonicity, leading to a shift of water out of cells. This shift of water is also associated with a shift of potassium out of cells into the extracellular space. Potassium shifts are further enhanced by the presence of acidosis and the breakdown of intracellular protein secondary to insulin deficiency. Furthermore, entry of potassium into cells is impaired in the presence of insulinopenia (Figure 54.2) [2].

Increasing evidence indicates that hyperglycemia in patients with DKA is associated with a severe inflammatory state characterized by an elevation of proinflammatory cytokines, reactive oxygen species, and cardiovascular risk factors in the absence of obvious infection or cardiovascular pathology which is improved by the anti-inflammatory effect of insulin and/or the improvement of the stress phenomena brought on by hyperglycemia and hyperlipidemia [20]. Circulating levels of tumor necrosis factor-α, interleukin [IL]-6, IL1-β, and IL-8, C-reactive protein, plasminogen activator inhibitor-1, FFA, cortisol, and growth hormone are significantly increased two- to fourfold on admission in patients with hyperglycemic crises compared with control subjects, and levels returned to normal after insulin treatment and resolution of hyperglycemic crises. Table 54.3 depicts the levels of proinflammatory cytokines, cardiovascular risk factors, counterregulatory hormones, lipid peroxidation and the state of reactive oxygen species in obese and lean patients in hyperglycemic crises at admission and recovery versus obese and lean controls [20].

**Diagnosis**

**History and physical examination**

Clinical features of DKA at presentation can be nonspecific but, in general, patients complain of polydypsia and polyuria prior to a period of dehydration. These symptoms are often accompanied by nausea, vomiting, dizziness, and a rapid rise in body temperature. The patient may also experience watery diarrhea or increased urine output. The skin may appear flushed, and the patient may have shallow respiration. Physical examination may reveal tachycardia, hypertension, and oliguria. laboratory findings may include a urinary glucose concentration of greater than 20 g/L and a urine osmolality greater than or equal to 350 mOsm/kg. Additionally, the serum osmolality may be greater than 320 mOsm/kg, and the serum sodium may be greater than 150 mEq/L. The serum potassium may be lower than normal, and the serum bicarbonate concentration is often decreased. A blood glucose level of greater than 200 mg/dL is diagnostic of DKA.}

---

**Figure 54.2** Biochemical changes that occur during DKA. ATP, adenosine triphosphate; CoA, acetyl CoA; F 1,6, fructose 1,6; G-1-P, glucose-1-phosphatase; G-6-P, glucose-6-phosphatase; GH, growth hormone; HK, hexokinase; HMP, hexose monophosphate shunt; PEP, phosphoenolpyruvate; PFK, phosphofructokinase; PK, pyruvate kinase; TCA, tricarboxylic acid cycle; TG, triglyceride. Source: Adapted from Kitabchi et al. 2001 [10].
to the development of DKA. Generalized weakness, weight loss, and gastrointestinal symptoms including nausea, vomiting, and abdominal pain are frequently present on admission [21]. A prospective study of 189 consecutive patients with DKA reported that abdominal pain was present in 46% of patients with DKA and its presence related to the severity of metabolic acidosis and not to the severity of hyperglycemia or dehydration [21]. In DKA subjects with abdominal pain, the mean serum bicarbonate (9 ± 1 mmol L\(^{-1}\)) and blood pH (7.12 ± 0.02) were lower than in patients without pain (15 ± 1 mmol L\(^{-1}\) and 7.24 ± 0.09, respectively). Delayed gastric emptying and ileus induced by electrolyte disturbance and metabolic acidosis have also been implicated as possible causes of abdominal pain in DKA [22]. In the majority of patients, the abdominal pain spontaneously resolves after correction of the metabolic disturbance; thus, in the absence of an overt cause for abdominal pain, allowing several hours to treat the underlying acidosis constitutes the best diagnostic tool to elucidate the etiology of abdominal pain in DKA.

Physical examination reveals signs of dehydration, including loss of skin turgor, dry mucous membranes, tachycardia, and hypotension. Most patients are normothermic or even hypothermic at presentation, even in the presence of infections. Acetone on breath and labored Kussmaul respiration may also be present on admission, particularly in patients with severe metabolic acidosis. Mental status can vary from full alertness to profound lethargy; however, fewer than 20% of patients are hospitalized with loss of consciousness [23,24]. Abnormalities in mental status has been found to be independently associated with the level of acidosis as measured by arterial pH as compared to hyperosmolarity or serum “ketone” levels [25]. However, acidosis and hyperosmolarity work in synergy to produce depressed sensorium. Therefore, at presentation, a combination of hyperosmolarity and severe acidosis identifies high-risk adult patients who may benefit from more aggressive therapy and monitoring [25].

### Laboratory evaluation
The initial laboratory evaluation should include determination of arterial blood gases, blood glucose, blood urinary nitrogen (BUN), serum electrolytes, creatinine, ketones, osmolality, and urinalysis as well as a complete blood count with differential and bacterial culture for various tissues, as indicated [10]. While waiting for the laboratory tests, an initial blood glucose test by finger stick as well as a dipstick urinalysis for glucose, ketone, nitrite, and leukocyte esterase can give the clinician an insight into the severity and precipitating factors leading to metabolic decompensation. Because of the rather restrictive criteria for the diagnosis of DKA in the past, the recent ADA guidelines have provided three classification levels of DKA to outline different approaches to therapy, depending on the degree of ketoacidosis (Table 54.1).

The plasma anion gap, which is not a criterion for diagnosis, is calculated by subtracting the major measured anions (chloride and bicarbonate) from the major measured cation (sodium). The normal anion gap has been historically reported to be 12 ± 2 mEq L\(^{-1}\) [2]. Although most subjects with DKA present with a high anion gap acidosis, some patients present with mixed acid–base disorders [26]. It has been reported that 46% of patients admitted with DKA had predominant anion gap acidosis, 43% had mixed anion gap acidosis and hyperchloremic metabolic acidosis, and 11% had hyperchloremic metabolic acidosis [24]. During treatment, a transient hyperchloremic normal anion gap metabolic acidosis has been reported in the majority of patients [10].

### Table 54.3 Proinflammatory cytokines, cardiovascular risk factors, counterregulatory hormones, lipid peroxidation (TBA), and DCF values on admission and resolution of hyperglycemic crises in lean and obese DKA and obese hyperglycemic patients, compared with lean and obese nondiabetic subjects

<table>
<thead>
<tr>
<th></th>
<th>Lean DKA</th>
<th>Obese DKA</th>
<th>Obese hyperglycemia</th>
<th>Lean control</th>
<th>Obese control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Adm</td>
<td>Resol</td>
<td>Adm</td>
<td>Resol</td>
<td>Adm</td>
</tr>
<tr>
<td><strong>TNF-α (pg mL(^{-1}))</strong></td>
<td>22.7 ± 3.6</td>
<td>4.6 ± 0.9(^{a})</td>
<td>28.3 ± 2.8</td>
<td>5.9 ± 0.7(^{b})</td>
<td>24 ± 3.1</td>
</tr>
<tr>
<td><strong>IL-1β (pg mL(^{-1}))</strong></td>
<td>9.8 ± 2.3</td>
<td>1 ± 0.2(^{a})</td>
<td>13.7 ± 2.1</td>
<td>2.4 ± 0.3(^{b})</td>
<td>11 ± 0.8</td>
</tr>
<tr>
<td><strong>IL-6 (pg mL(^{-1}))</strong></td>
<td>14.9 ± 2.6</td>
<td>3.9 ± 1.1(^{a})</td>
<td>12.6 ± 2.1</td>
<td>4.3 ± 0.6(^{b})</td>
<td>10 ± 1.7</td>
</tr>
<tr>
<td><strong>IL-8 (pg mL(^{-1}))</strong></td>
<td>29.3 ± 3.4</td>
<td>10.6 ± 2.3(^{a})</td>
<td>27.4 ± 3.6</td>
<td>12 ± 2.8(^{b})</td>
<td>26 ± 3.4</td>
</tr>
<tr>
<td><strong>CRP (mg L(^{-1}))</strong></td>
<td>51 ± 3</td>
<td>28 ± 1.4(^{a})</td>
<td>59 ± 13</td>
<td>34 ± 9(^{b})</td>
<td>28 ± 6</td>
</tr>
<tr>
<td><strong>Homocysteine (μM)</strong></td>
<td>4.7 ± 0.2</td>
<td>3.7 ± 0.2(^{a})</td>
<td>5.9 ± 0.9</td>
<td>5.4 ± 0.7</td>
<td>3.7 ± 0.4</td>
</tr>
<tr>
<td><strong>FFA (mM)</strong></td>
<td>1.6 ± 0.1</td>
<td>0.5 ± 0.1(^{a})</td>
<td>1.4 ± 0.1</td>
<td>0.7 ± 0.1(^{a})</td>
<td>1.2 ± 0.2</td>
</tr>
<tr>
<td><strong>DCF (μM)</strong></td>
<td>8.6 ± 0.8</td>
<td>3.7 ± 0.5(^{a})</td>
<td>8.9 ± 1.2</td>
<td>4.1 ± 0.7(^{a})</td>
<td>7.8 ± 0.6</td>
</tr>
<tr>
<td><strong>TBA (μM)</strong></td>
<td>3.8 ± 0.7</td>
<td>1.3 ± 0.4(^{a})</td>
<td>4.0 ± 0.6</td>
<td>1.6 ± 0.2(^{a})</td>
<td>3.3 ± 0.5</td>
</tr>
<tr>
<td><strong>PAI-1 (ng mL(^{-1}))</strong></td>
<td>42.1 ± 12.2</td>
<td>4.2 ± 2.1(^{a})</td>
<td>40.4 ± 12.4</td>
<td>13.0 ± 3.4(^{b})</td>
<td>35.4 ± 9.3</td>
</tr>
<tr>
<td><strong>GH (ng mL(^{-1}))</strong></td>
<td>12.3 ± 2.2</td>
<td>3.2 ± 1.0(^{a})</td>
<td>10.0 ± 3.1</td>
<td>4.0 ± 1.2(^{a})</td>
<td>1.6 ± 0.3(^{a})</td>
</tr>
<tr>
<td><strong>Cortisol (μg dl(^{-1}))</strong></td>
<td>46.2 ± 2.3</td>
<td>21.7 ± 1.1(^{a})</td>
<td>55.4 ± 5.8</td>
<td>24.6 ± 3.6(^{b})</td>
<td>23.0 ± 9(^{a})</td>
</tr>
</tbody>
</table>

Data are mean ± SE.

Adm, admission; CRP, C-reactive protein; FFA, free fatty acid; PAI-1, plasminogen activator inhibitor-1; Resol, resolution.

\(^{a}\)p < 0.01 vs. lean DKA on admission.

\(^{b}\)p < 0.05 vs. admission value of each group.
Effective osmolality can be calculated by the following formula: 2(measured Na+ (mEq·L⁻¹) + glucose (mg·dL⁻¹))/18. The sodium concentration in DKA is not adjusted to the glucose concentration for calculation of osmolality. However, the sodium concentration should be adjusted by glucose concentration for clinical treatment decisions when assessing actual serum sodium concentration. Table 54.4 outlines the laboratory evaluation of metabolic causes of acidosis and coma [10].

Laboratory pitfalls
Clinical assessment of ketonuria and ketonemia has traditionally been done by the use of the nitroprusside tablet or Ketostix, which only reacts with acetoacetate and acetone [10,27] and fails to detect β-hydroxybutyrate, the most important ketone body in patients with DKA. During therapy, β-hydroxybutyrate is converted to acetoacetate, so measurement of acetoacetate by the nitroprusside test as an indication of DKA recovery, may be misleading. The recent availability of self-glucose monitoring systems, which also contain a mechanism for measurement of β-hydroxybutyrate by finger stick, has led to a marked improvement in the management of patients with DKA [2,27,28].

Leukocytosis, which is common in patients with DKA, may not be indicative of an infectious process, particularly when total white blood cell counts are below 25,000 [29]. The reason for this elevation has not been established. The majority of white cells in leukocytosis are granulocytes with no detectable eosinophil counts (as a result of hypercortisolism) [29]. Leukocytosis (total white blood cell counts >25,000) requires further assessment for an underlying infection. The admission serum sodium is usually low because of the osmotic flux of water from the intracellular to the extracellular space in the presence of hyperglycemia. An increase in serum sodium concentration in the presence of hyperglycemia indicates a rather profound degree of water loss. To assess the severity of sodium and water deficit, serum sodium may be corrected by adding 1.6 mg·dL⁻¹ to the measured serum sodium for each 100 mg·dL⁻¹ of glucose above 100 mg·dL⁻¹. Extreme hypertriglycerideremia, which may be present during the DKA may cause lipemic serum with spurious lowering of serum sodium (pseudohyponatremia) [16]. Serum potassium is commonly elevated as a result of cellular dehydration and shift of potassium from the intracellular to the extracellular space [30], and to the breakdown of intracellular protein secondary to insulin deficiency.

Other laboratory abnormalities in patients with DKA include hyperamylasemia, which has been reported in 21–79% of cases [31]. There is little correlation between the presence, degree, or isoenzyme type of hyperamylasemia and the presence of gastrointestinal symptoms, triglyceride concentration, or pancreatic imaging studies. The cause of hyperamylasemia in DKA is postulated to be multifactorial; the most important causes include the existence of salivary amylase, reduced renal clearance of amylase, and increased leakage from acinar cells secondary to neural and metabolic disturbances. Non specific serum lipase elevation has also been noted in 29–41% of patients with DKA in the absence of clinical or radiologic evidence of acute pancreatitis.

Recent studies have shown decreased serum leptin in obese and thin patients with DKA as well as obese hyperglycemic patients. These levels normalize after correction of hyperglycemia [32]. Furthermore, recent studies on proinflammatory factors in hyperglycemic crises demonstrate significant increases in TNF-α, IL-1, IL-6, and IL-8 [20]. These increased cytokines are also associated with de novo emergence of growth factor in activated T-lymphocytes for insulin, IGF-1, and IL-2 [33].

Hyperglycemic hyperosmolar state
Definition
Hyperglycemic hyperosmolar syndrome (HHS) was first described by Dreschfeld [32] over a century ago, but this condition received little attention until 1957, when Sament and Schwartz [34] reported their experience of a diabetic syndrome characterized by marked stupor without ketosis (diabetic hyperosmolar syndrome). HHS is characterized by extreme elevations in serum glucose concentrations and hyperosmolality without significant ketosis may occur in patients with T1 and T2DM. Criteria for the diagnosis of HHS include a serum glucose concentration >600 mg·dL⁻¹ (33 mmol·L⁻¹), a serum osmolality >330 mOsm·kg⁻¹, and lack of significant ketosis and acidosis (serum bicarbonate >5 mmol·L⁻¹, serum pH >7.30) [2]. Overlap between HHS and DKA may occur, and some patients with HHS—especially when there is severe dehydration—experience mild or moderate acidosis (Table 54.1). Altered sensorium (lethargy, stupor, and coma) is common and correlates better with hyperosmolality than with patient’s age or severity of acid–base disturbance. Several reports have shown that the mean total serum osmolality in patients who presented in coma is greater than 340 mmol·kg⁻¹ [35,36].

Epidemiology
Although epidemiologic reports on HHS are sparse, data collected from 1979 to 1981 by an NIH data group suggest an incidence of 10 cases per 100,000 of the general population [1]. This is about one sixth that of DKA. More recently, the incidence of HHS in two different teaching institutions in the United States was 0.05% of all diabetes-related admissions [37]. Mortality attributed to HHS is considerably higher than with DKA, with recent mortality rates of 5–35%, most likely depending on the underlying illness or comorbidities [35].

Precipitating and predisposing factors
In general, subjects with HHS have a history of symptoms of hyperglycemia for several days and patients present to the emergency room with an altered mental status [38]. The typical patient with HHS is between 55 and 70 years old and frequently
# Table 54.4 Laboratory evaluation of metabolic causes of acidosis and coma

<table>
<thead>
<tr>
<th>Condition</th>
<th>pH</th>
<th>Plasma glucose</th>
<th>Glycosuria</th>
<th>Total plasma ketones</th>
<th>Anion gap</th>
<th>Osmolality</th>
<th>Uric acid</th>
<th>Miscellaneous</th>
</tr>
</thead>
<tbody>
<tr>
<td>Starvation or high fat intake</td>
<td>Normal</td>
<td>Normal</td>
<td>Negative</td>
<td>Slight</td>
<td>Slight</td>
<td>Normal</td>
<td>Normal</td>
<td>May give false-positive for ethylene glycol</td>
</tr>
<tr>
<td>DKA</td>
<td>Mild</td>
<td>Normal</td>
<td>Negative</td>
<td>Normal</td>
<td>Slight</td>
<td>Normal</td>
<td>Normal</td>
<td>Serum lactate &gt;7mmol/L</td>
</tr>
<tr>
<td>Lactic Acid</td>
<td></td>
<td>Normal</td>
<td>Negative</td>
<td>Normal</td>
<td>Slight to moderate</td>
<td>Normal</td>
<td>Normal</td>
<td>Serum salicylate levels positive</td>
</tr>
<tr>
<td>Uremic acidosis</td>
<td></td>
<td>Normal</td>
<td>Negative</td>
<td>Normal</td>
<td></td>
<td>Normal</td>
<td>Normal</td>
<td>Serum salicylate positive</td>
</tr>
<tr>
<td>Alcoholic ketosis (starvation)</td>
<td></td>
<td>Normal</td>
<td>Negative</td>
<td>Normal</td>
<td></td>
<td>Normal</td>
<td>Normal</td>
<td>Serum salicylate levels positive</td>
</tr>
<tr>
<td>Salicylate Intoxication</td>
<td></td>
<td>Normal</td>
<td>Negative</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Serum salicylate levels positive</td>
</tr>
<tr>
<td>Methanol or ethylene glycol</td>
<td>Normal</td>
<td>&gt;500 mg/dl</td>
<td>&lt;30 mg/dl</td>
<td>Normal or slight</td>
<td>Normal or slight</td>
<td>Normal or slight</td>
<td>Normal</td>
<td>Serum salicylate levels positive</td>
</tr>
<tr>
<td>intoxication</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Normal</td>
<td>Serum salicylate levels positive</td>
</tr>
<tr>
<td>Hypersmolar coma</td>
<td>Normal</td>
<td>Normal</td>
<td>Negative</td>
<td>Normal or slight</td>
<td>Normal or slight</td>
<td>Normal or slight</td>
<td>Normal</td>
<td>Serum salicylate levels positive</td>
</tr>
<tr>
<td>Hypoglycemic coma</td>
<td>Normal</td>
<td>Normal</td>
<td>Negative</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Serum salicylate levels positive</td>
</tr>
<tr>
<td>Rhabdomyolysis</td>
<td>Normal</td>
<td>Normal</td>
<td>Negative</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Serum salicylate levels positive</td>
</tr>
</tbody>
</table>

Note: +, positive

1. Acetest and Ketostix measure acetoacetic acid only, thus, misleading low values may be obtained because the majority of “ketone bodies” are β-hydroxybutyrate.
2. Respiratory alkalosis/metabolic acidosis.
3. May get false-positive or false-negative urinary glucose caused by the presence of salicylate or its metabolites.

(Reproduced with permission of the American Diabetes Association)

is a nursing home resident or someone who has delayed seeking medical attention [36,38]. Infection is the major precipitating factor occurring in 30–60% of patients with, urinary tract infections and pneumonia being the most common infections [36,38,39].

HHS is the initial manifestation of diabetes in 7–17% of patients [36–39]. Other precipitating causes of HHS include an acute illness, such as cerebrovascular accident or myocardial infarction that provokes the release of counterregulatory hormones and/or compromises access to water [36–39]. Certain medications that cause DKA may also precipitate the development of HHS, including glucocorticoid, thiazide diuretics, dilantin, and β-blockers. See Table 54.2 for predisposing or precipitating factors for HHS.

Pathogenesis

The underlying pathogenic mechanism for HHS is a reduction in the net effective concentration of circulating insulin and a concomitant elevation of counterregulatory or stress hormones such as glucagon, catecholamines, cortisol, and growth hormone in association with severe dehydration [35,36]. DKA and HHS represent opposite ends of the spectrum of hyperglycemia with and without ketosis, respectively (Figure 54.2). However, mixed mild ketotic hyperosmolar conditions may exist, with about one third of DKA patients having features of HHS [40].

Several mechanisms have been proposed to explain the absence or minimal presence of ketosis in patients with HHS. Key differences between DKA and HHS include: (a) higher levels of endogenous insulin reserve in HHS (i.e., adequate insulin concentration to prevent lipolysis but inadequate to inhibit hepatic glucose production and/or stimulate glucose use); (b) lower levels of counterregulatory hormones and FFA in HHS; and (c) inhibition of lipolysis by the hyperosmolar state, thereby decreasing ketogenesis. Early studies comparing 12 cases of HHS and 22 cases of DKA confirmed a higher level of growth hormone, but revealed no significant difference in either FFA or cortisol in the two groups (Table 54.5) [41]. Although earlier studies were not able to demonstrate differences in pancreatic insulin secretion between patients with HHS and DKA, it is important to note that prior to 1980, the measurement of insulin was not specific and could not separate proinsulin from insulin and a standardized C-peptide assay was not used. More recent studies [20] have demonstrated a higher level of C-peptide in HHS than in DKA [20]. Similar results were observed recently in lean African-American patients with DKA as compared to obese DKA and obese nonketotic hyperglycemia state (Table 54.6) [20]. Therefore, the most likely reason for the lack of ketoadiposis in HHS appears to be a higher level of pancreatic insulin reserve than in DKA. This is due to the fact that the antilipolytic effect of insulin is about one tenth of that of glucose utilization [10]. Another potential mechanism for the lack of ketosis in HHS involves the effect of hyperosmolality on inhibiting lipolysis, insulin secretion, and glucose uptake. Hepatic glucose production in HHS is higher than that in DKA and may be as high as 1 kg per day. Equally important is the glomerular filtration rate, which may be decreased with increasing age, exacerbating the level of hyperglycemia, decreasing intravascular space, and leading to tachycardia and hypotension.

Diagnosis

As shown in Table 54.1, the diagnosis of HHS is characterized by a higher level of glucose levels, absence of metabolic acidosis, minimal or negative ketosis, and higher serum osmolality [10]. These patients are obtunded more often than DKA patients. Furthermore, biochemical data on admission suggest slightly higher sodium, BUN, creatinine, osmolality, and possibly higher serum lactate levels than in DKA (Table 54.5) [41].

The typical patient with HHS has undiagnosed diabetes, is between 55 and 70 years of age, and frequently is a nursing home resident. Most patients who develop HHS do so over days to weeks during which they experience polyuria, polydipsia, and progressive decline in the level of consciousness. The most common clinical presentation for patients with HHS is altered sensorium [39]. Physical examination reveals signs of volume depletion. Fever due to underlying infection is common, and signs of acidosis (Kussmaul respiration, acetone breath) are usually absent. Gastrointestinal manifestations (abdominal pain, vomiting) frequently reported in patients with DKA are not part of HHS, thus the presence of abdominal pain in patients without significant metabolic acidosis needs to be investigated. In some patients, focal neurologic signs (hemiparesis, hemianopsia) and seizures (partial motor seizures more common than generalized) may be the dominant clinical features, resulting in a common misdiagnosis of stroke [39].
Despite the focal nature of neurologic findings, the neurologic manifestations often reverse completely after correction of metabolic disorder.

The diagnostic criteria for HHS include a plasma glucose concentration greater than 600 mg dL\(^{-1}\), a serum osmolality greater than 320 mOsm kg\(^{-1}\) of water, and the absence of significant ketoadiposis. Although by definition, patients with HHS have a serum pH greater than 7.3, a serum bicarbonate greater than 18 mEq L\(^{-1}\), and negative ketone bodies in urine and plasma, mild ketonemia may be present. Approximately 50% of the patients with HHS have an increased anion gap metabolic acidosis as the result of concomitant ketoadiposis and/or an increase in serum lactate levels [5].

### Treatment of hyperglycemic crises

Because DKA and HHS have essentially the same underlying pathogenesis consisting of insufficient insulin and loss of fluid and electrolytes, the major therapeutic goals include: (a) improvement of organ perfusion by increasing circulatory volume, (b) gradual reduction of osmolality, (c) gradual correction of serum glucose, (d) clearance of both serum and urine of ketones, and (e) normalization of electrolyte levels [2,16]. Careful and frequent monitoring of these patients every 30–60 min for the first 4 hours and subsequently every hour is the most important aspect of management of patients with hyperglycemic crises. Figure 54.3 summarizes the algorithm for the treatment of DKA and HHS [42].

### Hydration and fluid therapy

As reported by previous investigators, hydration alone has been shown to increase extra cellular volume, lower serum glucose, decrease hyperosmolality, reduce circulating levels of counterregulatory hormones, and improve insulin resistance [43]. Because of severe loss of intravascular volume and electrolytes in both DKA and HHS, the initial fluid therapy should be one liter of isotonic saline, which will expand extracellular fluid and restore renal perfusion. The rate of isotonic solution should be given at 15–20 mL kg\(^{-1}\) h\(^{-1}\). The tonicity and volume of subsequent sodium chloride solutions should be determined by the state of hydration and by the corrected serum sodium value (Figure 54.3). In general, 0.45% sodium chloride solution after the initial normal saline will expand intracellular and interstitial space volume and normalize the contracted volume within a few hours, and should continue until blood glucose reaches 200 mg dL\(^{-1}\) in DKA or HHS. At this time, the hydrating fluid should be changed to 5–10% glucose-containing saline with a concomitant reduction of i.v. insulin to 0.02–0.05 U kg\(^{-1}\) h\(^{-1}\). Plasma glucose levels should remain ~200 mg dL\(^{-1}\) until bicarbonate and pH reach normal levels in DKA and serum osmolality and mental status in patients with HHS is normalized. An additional important aspect of fluid management in hyperglycemic states is to replace the volume of urinary losses. Failure to adjust fluid replacement for urinary losses may delay correction of electrolytes and water deficit [10].

### Insulin therapy

The cornerstone of DKA and HHS management is insulin therapy. Insulin increases peripheral glucose utilization and decreases hepatic glucose production, thereby lowering blood glucose concentration. In addition, insulin therapy inhibits the release of FFA from adipose tissue and decreases ketogenesis, both of which lead to the reversal of ketogenesis. In critically ill and mentally obtunded patients, regular insulin given intravenously by continuous infusion is the treatment of choice. Such patients should be admitted to an intensive care unit or to a step-down unit where adequate nursing care and quick

### Table 54.6 Clinical characteristics and pancreatic C-peptide of DKA and hyperglycemia subjects vs. controls

<table>
<thead>
<tr>
<th></th>
<th>Lean DKA</th>
<th>Obese DKA</th>
<th>Obese hyperglycemia</th>
<th>Obese control</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>54</td>
<td>77</td>
<td>51</td>
<td>25</td>
</tr>
<tr>
<td>Age (years)</td>
<td>36 ± 2</td>
<td>40 ± 2</td>
<td>46 ± 2</td>
<td>38 ± 2</td>
</tr>
<tr>
<td>Sex (MF)</td>
<td>34/20</td>
<td>49/28</td>
<td>34/17</td>
<td>9/16</td>
</tr>
<tr>
<td>New-onset diabetes (% of total)</td>
<td>15 (28)</td>
<td>64 (83)</td>
<td>40 (78)</td>
<td>2</td>
</tr>
<tr>
<td>BMI (kg/m(^2))</td>
<td>22 ± 1</td>
<td>37 ± 1</td>
<td>37 ± 1</td>
<td>30 ± 2</td>
</tr>
<tr>
<td>Glucose (mmol/l)</td>
<td>36 ± 2</td>
<td>36 ± 2</td>
<td>37 ± 2</td>
<td>5 ± 1</td>
</tr>
<tr>
<td>Venous pH</td>
<td>7.18 ± 0.02</td>
<td>7.24 ± 0.01</td>
<td>7.4 ± 0.01</td>
<td>—</td>
</tr>
<tr>
<td>Bicarbonate (mEq/l)</td>
<td>12 ± 1</td>
<td>14 ± 1</td>
<td>22 ± 1</td>
<td>24 ± 2</td>
</tr>
<tr>
<td>HbA(_1c) (%)</td>
<td>13.0 ± 0.4</td>
<td>13.5 ± 0.4</td>
<td>12.4 ± 0.4</td>
<td>5.4 ± 0.1</td>
</tr>
<tr>
<td>Basal C-peptide (ng/ml)</td>
<td>0.5 ± 0.1(^\dagger)</td>
<td>1.5 ± 0.1</td>
<td>1.8 ± 0.1</td>
<td>1.9 ± 0.1</td>
</tr>
<tr>
<td>Stimulated C-peptide(^\dagger)</td>
<td>0.6 ± 0.1(^\dagger)</td>
<td>2.5 ± 0.1</td>
<td>3.5 ± 0.3(^\dagger)</td>
<td>4.2 ± 0.3</td>
</tr>
<tr>
<td>Acute C-peptide response(^\dagger)</td>
<td>0.2 ± 0.13(^\dagger)</td>
<td>1.0 ± 0.1</td>
<td>1.7 ± 0.2(^\dagger)</td>
<td>2.3 ± 0.2</td>
</tr>
</tbody>
</table>

Note: Data are means ± SEM

\(^{\dagger}\)p < 0.01 vs. obese DKA

Diabetic ketoacidosis and hyperosmolar state

Cardiogenic shock

Check electrolytes, BUN, venous pH, creatinine and glucose every 2–4 hrs until stable.

After resolution of DKA or HHS and when patient is able to eat, initiate SC multidose insulin regimen.

To transfer from IV to SC, continue IV insulin infusion for 1–2 hr after SC insulin begun to ensure adequate plasma insulin levels. Keep serum glucose between 150 and 200 mg/dl until resolution of DKA/HHS.

If serum glucose does not fall by at least 10% in first hour, give 0.14 U/kg/B.Wt as IV bolus, then continue previous Rx.

When serum glucose reaches 200 mg/dl, reduce regular insulin infusion to 0.02–0.05 U/kg/hrIV.

When stable and mentally alert, initiate rapid-acting insulin at 0.1 U/kg SC every 2 hrs. To transfer from IV to SC insulin, continue IV insulin infusion for 1–2 hours after SC insulin begun to ensure adequate plasma insulin levels. Keep serum glucose between 150 and 200 mg/dl until resolution of DKA/HHS.

If serum glucose does not fall by at least 10% in first hour, give 0.14 U/kg/B.Wt as IV bolus, then continue previous Rx.

When serum glucose reaches 200 mg/dl, reduce regular insulin infusion to 0.02–0.05 U/kg/hrIV.

When stable and mentally alert, initiate rapid-acting insulin at 0.1 U/kg SC every 2 hrs. To transfer from IV to SC, continue IV insulin infusion for 1–2 hr after SC insulin begun to ensure adequate plasma insulin levels. Keep serum glucose between 150 and 200 mg/dl until resolution of DKA/HHS.

If serum glucose does not fall by at least 10% in first hour, give 0.14 U/kg/B.Wt as IV bolus, then continue previous Rx.

When serum glucose reaches 200 mg/dl, reduce regular insulin infusion to 0.02–0.05 U/kg/hrIV.

When stable and mentally alert, initiate rapid-acting insulin at 0.1 U/kg SC every 2 hrs. To transfer from IV to SC, continue IV insulin infusion for 1–2 hr after SC insulin begun to ensure adequate plasma insulin levels. Keep serum glucose between 150 and 200 mg/dl until resolution of DKA/HHS.

If serum glucose does not fall by at least 10% in first hour, give 0.14 U/kg/B.Wt as IV bolus, then continue previous Rx.

When serum glucose reaches 200 mg/dl, reduce regular insulin infusion to 0.02–0.05 U/kg/hrIV.

When stable and mentally alert, initiate rapid-acting insulin at 0.1 U/kg SC every 2 hrs. To transfer from IV to SC, continue IV insulin infusion for 1–2 hr after SC insulin begun to ensure adequate plasma insulin levels. Keep serum glucose between 150 and 200 mg/dl until resolution of DKA/HHS.

Figure 54.3 Protocol for management of adult patients with DKA or HHS.† Initial evaluation: After history and physical examination, obtain arterial blood gases, complete blood count with differential, urinalysis, plasma glucose, BUN, electrolytes, chemistry profile, and creatinine levels immediately; as well as ECG. Chest radiography and cultures as needed. Start i.v. fluids: 1.0 L of 0.9% NaCl per hour.† Source: Kitabchi AE, et al. Hyperglycemic crises in adult patients with diabetes. Diabetes Care 2009;32(7):1335–1343. Reproduced with permission of the American Diabetes Association.

*DKA diagnostic criteria: blood glucose 250 mg/dl, arterial pH 7.3, bicarbonate 15, mEq/L, and moderate ketonuria or ketonemia
† 15–20 mEq/kg
‡ Serum Na should be corrected for hyperglycemia (for each 100 mg/dl glucose 100 mg/dl, add 1.6 mEq to sodium value for corrected serum value)
Serial measurements of plasma glucose are usually sufficient to achieve and maintain ketone levels below 0.5 mmol/L. When plasma glucose levels reach 5–10 mmol/L, the insulin infusion rate is reduced to 0.02–0.05 U/kg·h (3–5 U/h), and dextrose (5–10%) should be added to i.v. fluids. Thereafter, the rate of insulin administration may need to be adjusted to maintain the above glucose values until ketoacidosis or mental obtundation and hyperosmolality are resolved. During therapy, capillary blood glucose should be determined every 1–2 hours at the bedside using a glucose oxidase reagent strip; and blood should be drawn every 2–4 hours for determination of serum electrolytes, glucose, blood urea nitrogen, creatinine, magnesium, phosphorus, and venous pH.

A conscious patient with mild DKA could be treated with subcutaneous (s.c.) or intramuscular (i.m.) insulin in a general hospital ward. In such patients, the administration of regular insulin every 1–2 hours by the s.c. or i.m. route has been shown to be as effective in lowering blood glucose and ketone bodies concentration as giving the entire insulin dose by i.v. infusion without addition of albumin [37,38]. Blood insulin levels using the i.m. and s.c. route may take 1–2 hours before reaching postprandial physiologic levels (70–100 μU/mL), whereas i.v. insulin reaches supraphysiologic insulin level in a few minutes. The use of i.v. and s.c. (or i.m.) insulin is an acceptable therapeutic approach in a non-ICU setting [37,38,39]. Recent prospective randomized studies comparing the use of rapid-acting insulin (lispro) s.c. route versus i.v. infusion in the ICU, in mild to moderate DKA patients, has been reported to be as effective as a standard i.v. low-dose regular insulin protocol. In these studies, the mean time of insulin treatment to correct hyperglycemia and ketoacidosis was similar between s.c. lispro and i.v. infusion of regular insulin [37].

**Transition to subcutaneous insulin**

Patients with DKA should be treated with continuous i.v. insulin until ketoacidosis is resolved. Criteria for resolution of ketoacidosis include: a blood glucose lower than 200 mg dL−1, a serum bicarbonate level equal to or greater than 18 mEq L−1, a venous pH greater than 7.3, and a calculated anion gap equal to or lower than 14 mEq L−1 [2]. Serial measurements of point-of-care capillary blood β-hydroxybutyrate (β-OHB) concentrations have been proposed to monitor biochemical resolution of DKA and to determine when i.v. insulin infusion should be stopped. Blood β-OHB levels decrease to <1 mmol L−1 many hours before the urine becomes ketone-free [44]. Absence of ketonuria should not be used as an endpoint for determining resolution of DKA.

The American Diabetes Association (ADA) position statement recommends the transition to a split-mixed insulin regimen with neutral protamine Hagedorn (NPH) and regular insulin twice daily or to a multidose regimen of short- or rapid-acting and intermediate- or long-acting insulin [2]. Several studies have reported hospital rates of hypoglycemic events up to 30% with the use of NPH and regular insulin after discontinuation of i.v. insulin [45]. The inadequate duration of action of NPH insulin and an undesirable peak activity at 4–6 hours after injection [45] as well as the high day-to-day variability in absorption [45] partially explain the high rate of hypoglycemic events. A recent randomized study compared the safety and efficacy of insulin analogues and human insulin during the transition from i.v. to s.c. insulin in patients with DKA. During the transition to s.c. insulin, there were no differences in mean daily glucose levels, but 41% patients treated with NPH/regular had a higher rate of hypoglycemia compared to 15% of patients treated with glargine once daily and glulisine before meals [45]. Thus, a basal/bolus regimen with insulin analogues is safer and should be preferred over NPH and regular insulin following the resolution of DKA. To prevent recurrence of hyperglycemia and/or ketoacidosis during the transition period to s.c. insulin, it is important to allow an overlap of 1–2 hours between discontinuation of i.v. insulin and the administration of s.c. insulin [45].

The most convenient time to change to s.c. insulin is in the morning or evening. For patients previously on a basal-bolus insulin regimen, the first s.c. injection of rapid-acting insulin should be given 15 min prior to the meal depending on the plasma glucose concentration. This method allows sufficient time for the injected insulin to be absorbed. Then the first basal injection is given and the insulin infusion is stopped 1 hour later.

In patients with established diabetes, the patient’s usual insulin regimen may be resumed. In adult patients with newly diagnosed diabetes, an initial insulin total insulin dose of 0.5–0.8 U kg−1 d−1 is usually sufficient to achieve and maintain metabolic control [2,37]. After transitioning to s.c. insulin, frequent blood glucose monitoring is required to avoid marked hyperglycemia and hypoglycemia. Supplemental rapid-acting insulin is given before meals or at 4–6-hour intervals to correct hyperglycemia.

**Potassium**

Despite a total body potassium deficit of ~3–5 mEq kg−1 of body weight, most patients with DKA have a serum potassium level at or above the upper limits of normal [46]. These high levels occur because of a shift of potassium from the intracellular to the extracellular space due to acidemia, insulin deficiency, and hypertonicity. Both insulin therapy and correction of acidosis decrease serum potassium levels by stimulating cellular potassium uptake in peripheral tissues. Therefore to prevent hypokalemia, most patients require i.v. potassium during the course of DKA therapy. Replacement with i.v. potassium should...
be initiated as soon as the serum potassium concentration is below 5.5 mEq L\(^{-1}\). The treatment goal is to maintain serum potassium levels within the normal range of 4 to 5 mEq L\(^{-1}\). In some hyperglycemic patients with severe potassium deficiency, insulin administration may precipitate profound hypokalemia [46], which can induce life-threatening arrhythmias and respiratory muscle weakness. Thus, if the initial serum potassium is lower than 3.3 mEq L\(^{-1}\), potassium replacement should begin immediately by an infusion of potassium chloride at a rate of 20–30 mEq h\(^{-1}\), and insulin therapy should be delayed in patients with significant hypokalemia (Figure 54.3).

**Bicarbonate**

Most current ADA guidelines [2] do not recommend the routine use of alkali therapy in DKA because bicarbonate levels tend to correct with insulin therapy. Insulin administration inhibits ongoing lipolysis and ketogenic production and promotes ketoanion metabolism. Because protons are consumed during ketoanion metabolism, bicarbonate is regenerated, leading to partial correction of metabolic acidosis. The use of bicarbonate in HHS is not recommended, as patients exhibit normal serum bicarbonate, unless there is a condition of mixed hyperosmolar ketoacidosis. The use of bicarbonate in DKA, however, may be used in patients with either severe bicarbonate deficiency (below 5 mM) or severe acidemia (pH less than 6.9). However, the use of bicarbonate in DKA at this level of acidosis is controversial because of the potential deleterious effects of bicarbonate including paradoxical central spine fluid (CSF) acidosis, rebound alkalaemia, hypokalemia, and increased lactate production as well as local irritation with bicarbonate administration. The theoretical advantage of bicarbonate therapy may be due to its salutary effect on advanced acidosis and improvement of the negative ionotropic effect on the heart and peripheral vasodilatation [47]. Several prospective and retrospective studies have demonstrated conflicting results regarding the use of bicarbonate therapy in DKA. Two studies have shown increases in blood pH when bicarbonate therapy was compared to saline infusion [47]. There are no large studies to demonstrate the efficacy of bicarbonate therapy in patients with DKA and pH less than 6.9. In the absence of such data, the ADA guidelines recommend that patients with pH lower than 6.9 should receive two ampules of bicarbonate (100 mmol) as an isotonic solution (in 400 mL of water) at the rate of 50 mmol h\(^{-1}\) with 20 mEq of KCl [2]. For follow-up of acidosis, we recommend measuring venous rather than arterial pH every 2 hours, realizing that venous pH is approximately 0.03 lower than the arterial pH.

**Phosphate**

Total body phosphate deficiency is universally present in patients with DKA, but its clinical relevance and benefits of replacement therapy remain uncertain. Several studies have failed to show any beneficial effect of phosphate replacement on clinical outcome [48]. Furthermore, aggressive phosphate therapy is potentially hazardous, as indicated in case reports of children with DKA who developed hypocalcemia and tetany secondary to i.v. phosphate administration [48]. Theoretic advantages of phosphate therapy include prevention of respiratory depression and generation of erythrocyte 2,3-diphosphoglycerate. Because of these potential benefits, careful phosphate replacement may be indicated in patients with cardiac dysfunction, anemia, respiratory depression, and in those with serum phosphate concentration lower than 1.0–1.5 mg dL\(^{-1}\) [48]. If phosphate replacement is needed, it should be administered as a potassium salt, by giving half of the potassium replacement as potassium phosphate and half as potassium chloride. In such patients, because of the risk of hypocalcemia, serum calcium and phosphate levels must be monitored during phosphate infusion.

**Management of patients after resolution of hyperglycemic crisis**

Patients with moderate to severe DKA should be treated with continuous i.v. insulin until ketoacidosis is resolved. Criteria for resolution of ketoacidosis include a blood glucose lower than 200 mg dL\(^{-1}\), a serum bicarbonate level equal to or greater than 18 mEq L\(^{-1}\), a venous pH greater than 7.3, and a calculated anion gap equal to or lower than 12 mEq L\(^{-1}\). The criteria for resolution of HHS include improvement of mental status, blood glucose <200 mg dL\(^{-1}\), and a serum osmolality of less than 320 mOsm kg\(^{-1}\). When this occurs, s.c. insulin therapy can be started. If the patient is able to eat, split-dose therapy with both rapid/short-acting insulin and basal insulin may be given. It is easier to make this transition in the morning—before breakfast—or at dinnertime. Patients with known diabetes may be given insulin at the dosage they were receiving before the onset of DKA. In patients with newly diagnosed diabetes, an initial insulin total insulin dose of 0.5–0.8 U kg\(^{-1}\) d\(^{-1}\) is usually sufficient to achieve and maintain metabolic control. Two thirds of this total daily dose should be given in the morning and one third in the evening as a split-mixed dose. If the patient is not able to eat, continue the i.v. insulin infusion protocol. However, the patient could receive s.c. regular insulin every 4 hours while an infusion of 5% dextrose in half-normal saline is given at a rate of 100–200 mL h\(^{-1}\). A critical element is to avoid recurrence of hyperglycemia or ketoacidosis during the transition period to s.c. insulin by allowing an overlap of i.v. insulin for 1–2 hours during the initiation of s.c. regular insulin to insure adequate plasma insulin levels.

Figure 54.3 illustrates step-by-step guidelines for management of patients with DKA and HHS approved by the ADA. It is important to point out, however, that no set of guidelines can replace individualized management of the patient in DKA or HHS.
Complications of therapy

The most common complications of DKA treatment are hypoglycemia and hypokalemia. Despite the use of low-dose insulin protocols, hypoglycemia is still reported in 10–25% of patients with DKA [2,37,49]. The failure to reduce insulin infusion rate and/or to use dextrose-containing solutions when blood glucose levels reach 250 mg dL\(^{-1}\) are the most important risk factors associated with hypoglycemia during insulin infusion. Frequent blood glucose monitoring (every 1–2 hours) is mandatory to recognize hypoglycemia and serious complications. Many patients with hyperglycemic crises who experience hypoglycemia during treatment do not experience adrenergic manifestations of sweating, nervousness, fatigue, hunger, and tachycardia despite low blood glucose levels. Clinicians should be aware that recurrent episodes of hypoglycemia may be associated with a state of hypoglycemia unawareness (loss of perception of warning symptoms of developing hypoglycemia), which may complicate diabetes management after resolution of hyperglycemic crises. Hypoglycemia is not a major problem in patients with HHS. Glucose values lower than 60 mg dL\(^{-1}\) have been reported in less than 5% of HHS patients during i.v. insulin therapy [2,37].

Although the admission serum potassium concentration is commonly elevated in patients with DKA and HHS, plasma concentration of potassium will invariably decrease as a result of treatment. Both insulin therapy and correction of acidosis decrease serum potassium levels by stimulating cellular potassium uptake in peripheral tissues. Thus, to prevent hypokalemia, replacement with i.v. potassium should be initiated when the serum potassium concentration is below 5.5 mEq L\(^{-1}\). In patients admitted with normal or reduced serum potassium, insulin administration may precipitate profound hypokalemia. Thus, if the initial serum potassium is less than 3.3 mEq L\(^{-1}\), i.v. potassium replacement should begin immediately and insulin therapy should be held until serum potassium reaches 3.3 mEq L\(^{-1}\) (see Figure 54.3).

Cerebral edema is a rare but serious complication of DKA, occurring in approximately 1% of episodes of diabetic ketoacidosis in children [44,45], and is associated with a mortality rate of 40–90% [50]. Clinically cerebral edema is characterized by decreasing level of consciousness and headache, followed by seizures, sphincter incontinence, pupillary changes, papilledema, bradycardia, and respiratory arrest. It has been hypothesized that cerebral edema in children with diabetic ketoacidosis may be caused by the rapid shift in extracellular and intracellular fluids and changes in osmolality due to accumulation of osmolytes in brain cells exposed to hyperosmolar conditions [50]. A rapid decrease in extracellular osmolality during treatment would then result in osmotically mediated swelling of the brain. Although osmotic factors and other mechanisms may play a part in the development of cerebral edema, recent data suggest that cerebral edema in children with diabetic ketoacidosis is related to brain ischemia [50]. In children with diabetic ketoacidosis, both hypocapnia (which causes cerebral vasoconstriction) and extreme dehydration (as determined by a high initial serum urea nitrogen concentration) were associated with increased risk for cerebral edema. Hyperglycemia superimposed on an ischemic insult increases the extent of neurologic damage, blood–brain barrier dysfunction, and edema formation.

In addition, it has been shown that an inadequate correction in serum sodium concentration during therapy may be associated with increased risk of cerebral edema [50]. The more frequent occurrence of cerebral edema in children than in adults may be explained in part by the fact that children's brains have higher oxygen requirements than adults' brains and are thus more susceptible to ischemia. Measures that may decrease the risk of cerebral edema in high-risk patients are gradual replacement of sodium and water deficits in patients with high serum osmolality (maximal reduction in osmolality 3 mOsm kg\(^{-1}\) h\(^{-1}\)) and the addition of dextrose to the hydrating solutions once blood glucose reaches 250 mg dL\(^{-1}\) in DKA and 300 mg dL\(^{-1}\) in HHS [10]. Patients with cerebral edema should be transferred to an intensive care unit setting. If signs of increased intracranial pressure or brain herniation are present, only 7–14% of patients recover without permanent significant neurologic disabilities. Treatment includes the immediate use of i.v. mannitol [10], reduction of fluid administration rate, and possible mechanical ventilation to help reduce brain swelling [10]. Corticosteroid and diuretic therapy are of no proven benefit in the immediate use of i.v. mannitol [10].

Prevention of hyperglycemic crises

One of the major precipitating factors in both DKA and HHS is inadequate insulin dosing or omission of insulin injection. The frequency of hospitalizations for DKA has been reduced following diabetes education programs, improved follow-up care, and access to medical advice. Outpatient management is more cost-effective and can minimize missed days of school and work for patients with diabetes and their family members. Important aspects of preventive care in HHS include frequent monitoring and adequate hydration as well as prevention and early treatment of hyperglycemia. Thus, diabetes education and sick-day management should be reviewed periodically in patients with diabetes and should include specific information on when to contact the healthcare provider, blood glucose and A1c goals, use of supplemental short- or rapid-acting insulin during illness, and, most important, the importance of never discontinuing insulin and of seeking immediate medical attention in the case of severe hyperglycemia. In addition, patients with T1DM should be instructed on the use of home blood ketone monitoring during illness and persistent hyperglycemia, which may allow for early recognition of impending ketoacidosis, and in turn may help to guide insulin therapy at home and, possibly, may prevent hospitalization for DKA. Finally, the alarming rise in insulin discontinuation because
of economic reasons as the precipitating cause for DKA in urban patients illustrates the need for healthcare legislation to ensure reimbursement for medications to treat diabetes. Novel approaches to patient education incorporating a variety of healthcare beliefs and socioeconomic issues are critical to an effective prevention program. Furthermore, recent studies in DKA suggest that frequent endocrine clinic visits on a quarterly rather than on an annual basis reduces DKA events significantly as compared to an annual visit [2].

Future prospects

The following are selected research topics and questions proposed to guide the management of patients with DKA and HHS in various hospital settings:
1. The mechanism for low or absent ketones in HHS versus DKA.
2. Left ventricular cardiac status of patients in severe DKA.
3. Resource utilization for the management of patients with mild, moderate, and severe DKA and HHS in various inpatient settings (ICU versus step-down units versus general wards).
4. What are optimal and safe glycemic targets during continuous insulin infusion until resolution of DKA? It is not known if we need to maintain a glucose concentration in the hyperglycemic range between \(\sim 150 \text{ mg dL}^{-1}\) and \(250 \text{ mg dL}^{-1}\) until resolution of ketoacidosis and HHS. This recommendation excludes the use of insulin infusion protocols commonly used in critically ill patients with hyperglycemia (glucose target \(140–180 \text{ mg dL}^{-1}\)).
5. Is the use of bicarbonate needed in patients with severe DKA? Available studies suggest that for pH >6.9–7.0, the use of bicarbonate does not provide any advantage. No prospective randomized studies are available to establish efficacy of the use of bicarbonate in DKA for pH <6.9.
6. What is the role of continuous glucose monitoring (CGM) systems in the acute and post discharge management of patients with hyperglycemic crises? The Endocrine Society Clinical Practice Guideline Committee on the use of CGM recently recommended against the use of this technology in critically ill patients. The use of CGM may have an advantage over bedside point-of-care (POC) testing in that it has the potential to reduce unknown hypoglycemic events that may occur between POC measurements; however, there is concern about the accuracy of CGM at low blood glucose levels.

Acknowledgments

The development of this document culminating 35 years of research and treatment of DKA and HHS could not have been possible without many contributors. Foremost among them have been more than 400 patients who so kindly agreed to participate in these studies. Other support was also provided by the Regional Medical Center in Memphis, TN and Grady Memorial Hospital in Atlanta, GA. The tremendous help of many nursing and technical staff of the General Clinical Research Center and the two hospitals are greatly appreciated. Last, but not least, the help and contributions of our colleagues at the institutions at Emory University (Atlanta, GA), the University of Washington (Seattle, WA), Virginia Mason Clinic (Seattle, WA), and the University of Tennessee College of Medicine (Memphis, TN) as well as more than 200 trainees and house staff of the Regional Medical Center and Grady Hospital have been immeasurable, without whom we could not have carried out these works successfully. The administrative assistance by Ms. Tara Bea is greatly appreciated.

References


SECTION X

Management of diabetes: glucose and lipid monitoring, specific subgroups
Type 2 diabetes in obese adolescents: pathophysiology and treatment

Grace Kim and Sonia Caprio
Department of Pediatrics, Yale University School of Medicine, New Haven, CT, USA

Key points
- The prevalence of type 2 diabetes (T2DM) is increasing in the pediatric population, particularly in ethnic minorities.
- Adolescents with impaired glucose tolerance have severe insulin resistance and abnormal insulin secretion.
- In comparison to adults, the time course in developing T2DM is more rapid in children and adolescents.
- Insulin resistance is the result of the imbalance of fat deposition in skeletal muscle and abdominal fat.
- Nonalcoholic fatty liver disease is a strong risk factor for T2DM.
- The treatment of T2DM in children and adolescents is a challenge and combination pharmacologic treatment for long-term glycemic control will need to be explored.

Type 2 diabetes in obese youth: a growing challenge

Type 2 diabetes mellitus (T2DM) has emerged as one of the greatest global health challenges of the twenty-first century, projected to affect roughly 1 in 3 individuals born in the year 2000 at some time during their lifetime [1]. Although T2DM was formerly a disease of adulthood, youth-onset T2DM now represents a substantial percentage of new cases of diabetes overall, ranging from 14% in non-Hispanic Whites to 86% in American Indians [2]. Recent population-based data from the SEARCH study (Search for Diabetes in Youth) indicate that diabetes is diagnosed in about 3700 obese youth annually in the USA [2].

Progression from normal glucose tolerance to overt T2DM involves an intermediate stage of hyperglycemia, characterized by impaired fasting glucose (IFG) and/or IGT, now known as prediabetes [3]. Recent reports have documented a high prevalence of prediabetes among children and adolescents [4]. In a multiethnic clinic-based population of obese children and obese adolescents, IGT was detected in 25% of the obese children and 21% of the obese adolescents, and silent T2DM was identified in 4% of the obese adolescents, irrespective of ethnicity [5]. This was the first study to highlight the high prevalence of prediabetes in the midst of the epidemic of childhood obesity [5]. The risk factors associated with IGT in that study were, in order of importance: (1) insulin resistance (estimated by the Homeostatic Model Assessment of Insulin Resistance (HOMA-IR), (2) fasting proinsulin, (3) 2-hour insulin level, and (4) fasting insulin [5].

Cruz et al. found that 28% of obese Hispanic children with a positive family history for T2DM had IGT, but found no cases of T2DM [6]. Gruters et al. reported that the prevalence of IGT was 36.3% among an obese multiethnic cohort of children and adolescents with a risk factor for T2DM [7]. High prevalence of IGT has also been reported in obese children from Thailand [8] and the Philippines [9]. Recent studies conducted in Italy in a large overweight and obese pediatric population reported a prevalence of prediabetes of 12% [10].

Clinical characteristic of T2DM in youth

The TODAY study (Treatment Options for type 2 Diabetes in Adolescents and Youth) the largest ethnically and geographically diverse group of pediatric subjects with T2DM ever assembled, has provided detailed information about the clinical characteristics of the disease in youth [11]. Key emerging important facts about youth-onset T2DM are: higher prevalence in females; presence of moderate to severe obesity; peak of onset at puberty; presence of acanthosis nigricans; greater prevalence among African Americans, Hispanic, and Native American children; strong family history of diabetes; and highly economically disadvantaged families [11]. Of note, 26% of TODAY’s participants presented at baseline with hypertension,
13% had microalbuminuria, 79% had a low high-density lipoprotein level, and 10% had hypertriglyceridemia [11].

**Beta-cell dysfunction: an early alteration in the development of T2DM**

Over the past decade several groups have studied the putative defects underlying the development of T2DM in youth. However, very few studies have described the natural history of the disease in childhood. In obese adolescents relevant information on β-cell function across the spectrum of glucose tolerance have been provided by several cross-sectional and longitudinal studies in which the state-of-the-art modeling of insulin secretion were applied to the OGTT and hyperglycemic clamp tests [12–14]. Employing the hyperglycemic clamp together with mathematical modeling of glucose-stimulated insulin secretion, Weiss et al. investigated the role of insulin secretion and proinsulin processing in glucose regulation in a group of 62 obese adolescents with different glucose tolerance status (30 NGT, 22 IGT, and 10 T2DM) [15]. This study showed that when compared with obese adolescents with a similar degree of insulin resistance, those with IGT and those with T2DM display a progressive loss of glucose sensitivity of β-cell first-phase secretion [12]. In contrast, the homeostatic loop between insulin sensitivity and glucose sensitivity of β-cell second-phase secretion is apparently preserved in IGT but disrupted in T2DM [15]. This is accompanied by alterations in the proinsulin-to-insulin processing, showing also a failure in the complex post-transcriptional event of insulin production [15]. Thus, IGT obese adolescents are characterized by an alteration of first-phase secretion, and several β-cell defects concur to cause diabetes in childhood [15]. Therefore, in a similar way to those observed in adult subjects [16], almost ~80% of the β-cell function is reduced or lost already at diagnosis [15].

In addition, differences in β-cell function have been described in various prediabetic conditions seen in obese adolescents, such as IFG or IGT or the combined IFG/IGT states [15]. Cali et al. documented that IFG, in obese adolescents, is primarily linked to alterations in glucose sensitivity of first-phase insulin secretion [17]. In contrast, adolescents with IGT are affected by a more severe degree of peripheral insulin resistance and reduction in first-phase secretion. Interestingly, the association between IFG and IGT resulted in a new additional defect in glucose sensitivity of second-phase insulin secretion and by a profound insulin resistance [17]. The role of “pre-existing” β-cell dysfunction risk in obese adolescents with normal glucose tolerance was subsequently revealed by Cali et al. in a longitudinal study in which β-cell function was assessed from OGTT using the Oral C-Peptide Minimal Model [13]. The disposition index (DI) is a measure for β-cell function relative to insulin sensitivity, and is calculated as the product of insulin sensitivity and β-cell function [18]. In adults, DI is a measure that is predictor for the development of T2DM [19]. In a longitudinal study, obese children who progressed to IGT had declining DI values over time compared to those whose glucose tolerance did not worsen (Figure 55.1(b)) [20]. In addition, adolescents who progressed to IGT had lower β-cell function at baseline than adolescents whose glucose tolerance did not worsen (Figure 55.1(a)) [20].

![Figure 55.1](image-url)
Natural history of IGT and the development of T2DM

In adults, the process of developing T2DM from a prediabetic state is gradual and may take place over 4–10 years [12]. In a longitudinal study, obese children and adolescents with IGT were followed for <2 years (21 months) [12]. Almost half of the subjects (45.5%) reverted to NGT, 30.3% remained IGT, and 24.2% progressed to diabetes [12]. This time course to developing diabetes was strikingly faster in children than adults [12]. Subjects that reverted to normal glucose tolerance weighed less at baseline and gained minimal weight compared to subjects that progressed to have diabetes [12]. In contrast, those who progressed to diabetes were heavier at baseline and gained a significant amount of weight, approximately 27 kg over the study period [12].

Insulin resistance and T2DM: fat accumulation in all the wrong tissues

The muscle

Defects in both insulin action and secretion are important in the pathogenesis of T2DM [21]. The exact timing and relative importance with which these two defects appear in the natural history of T2DM have been a matter of great debate [21]. To address these issues our group studied both cross-sectional and dynamic changes in insulin action and secretion in obese adolescents with IGT. The cross-sectional study revealed profound insulin resistance in the obese adolescents with IGT compared with those with NGT [22]. The insulin resistance was mainly accounted for by a reduction in nonoxidative glucose disposal (storage), since rates of glucose oxidation did not differ significantly between the groups [22]. Similar metabolic defects have been described in adults with overt T2DM [23]. Next to the disturbances in glucose metabolism, fat metabolism is also deranged in IGT subjects. In particular the lipid composition of the skeletal muscle tissue, where most (70%) of the whole glucose disposal occurs, has attracted much attention in the past decade as a major player in the development of muscle insulin resistance [24,25]. This area of investigation has been greatly advanced by the recent development and validation of spectroscopic techniques, which noninvasively and reliably quantify fat stored inside the myocytes; the intramyocellular fat (IMCL) [25]. Strong positive correlations between insulin resistance and IMCL have been reported in offspring of T2DM subjects, suggesting that a high IMCL content may be involved in the development of insulin resistance [26–28].

Using 1H-NMR in our metabolic studies to gather information on the muscle lipid composition we found an excessive accumulation of IMCL in the soleus muscle of IGT obese adolescents compared to age and adiposity matched NGT obese adolescents [22]. Using univariate analysis we found that IMCL content was inversely correlated with the glucose disposal ($r = -0.51; p < 0.01$) and with nonoxidative glucose metabolism in the two groups together ($r = -0.45; p < 0.03$) [22]. This study offered a novel insight into the pathogenesis of prediabetes in obese children and adolescents, namely that changes in glucose homeostasis are closely linked with altered partitioning of fat in skeletal muscle [22]. It is not entirely clear at the present time what leads to an overaccumulation of lipid in skeletal muscle tissues. Defects in muscle fatty acid oxidation and increased fatty acid flux to muscle have both been considered as potential players [24]. Triglyceride accumulation per se in muscle does not cause insulin resistance. Rather, overaccumulation of various lipid moieties in myocytes, such as long chain fatty acyl-CoA, have been shown to interfere directly with insulin signaling and glucose transport [29–32].

The abdominal fat: a complex layer of metabolically different adipose tissues

Visceral fat accumulation is known to be associated with features of the insulin resistance syndrome in adults and obese children [33], although the nature of this relationship and the relative importance of visceral versus subcutaneous abdominal fat remains a matter of debate. Altered distribution of fat between the subcutaneous and intra-abdominal compartment is associated with the development of IGT in obese adolescents [34]. The visceral-to-subcutaneous ratio was significantly greater in the IGT than in the NGT group. Both the enlarged visceral depot ($r = -0.63; p < 0.01$) and the visceral-to-subcutaneous ratio ($r = -0.66; p < 0.01$) were inversely related to the insulin-stimulated glucose metabolism [34]. These studies, therefore, suggested that insulin resistance is the result of the imbalance of fat distribution in both the abdominal adipose and skeletal muscle tissues. Furthermore, our data are consistent with results obtained in both human and animal models of lipodystrophy, in which there is an absence of subcutaneous adipose tissue, leading to increased circulation of lipids and their accumulation in both myocytes and the visceral depot [28,35]. Thus, it appears that the ability of peripheral subcutaneous fat tissue to vary its storage capacity is critical for regulating insulin sensitivity and ultimately protecting against diabetes. Consistent with this hypothesis is the evidence derived from the use of thiazolidinediones, which improve insulin sensitivity while increasing subcutaneous abdominal fat, but decreasing the visceral and liver fat depots [36,37].

The liver

Like T2DM, nonalcoholic fatty liver disease (NAFLD) is a comorbidity associated with obesity [38]. NAFLD is described as hepatic fat accumulation not caused by alcohol use [39]. NAFLD represents a spectrum that ranges from simple fatty deposition to inflammation, to fibrosis and cirrhosis [40]. Fatty liver may precede the onset of T2DM [38], and is a marker of insulin resistance and glucose dysregulation in children and adults [41]. Several adult studies have shown that high levels of hepatic enzymes (i.e., alanine aminotransferase (ALT))
forecast the development of T2DM [42]. In a multiethnic group of 392 adolescents, glucose and lipid metabolism measures were abnormal even at ALT levels in the upper half of normal range [41]. Using a noninvasive method, hepatic fat content was measured with fast magnetic resonance imaging, where HFF equal to or above 5.5% is suspicious for hepatic steatosis [41]. Thirty-two percent of subjects had increased hepatic fat fraction [41]. Hepatic steatosis was associated with decreased insulin sensitivity, and increased triglycerides and visceral fat (Figure 55.2) [41]. The connection between abdominal obesity and fatty liver may be due to excess free fatty acids travelling to the liver overflowing from the visceral fat compartment [33]. Subjects with fatty liver had a higher prevalence of metabolic syndrome [41].

Indeed, fatty liver is considered to be the hepatic component of metabolic syndrome [41]. In order to understand the potential role of fatty liver in the onset of T2DM in obese youth, we have recently assessed whether the severity of hepatic steatosis affects the presence of glucose metabolism dysregulation in a multiethnic cohort of obese adolescents [43]. Independent of obesity, the severity of fatty liver was associated with the presence of prediabetes conditions (IGT and IFG/IGT) and hepatic steatosis, independently, predicted prediabetes in obese adolescents [43]. In addition, paralleling the severity of hepatic steatosis, there was a significant decrease in insulin sensitivity and impairment of \( \beta \)-cell function (assessed by using the DI) [43]. These findings suggest that the intrahepatic fat accumulation is a strong risk factor for T2DM, and its early identification is critical to prevent the development of metabolic complications in youth [43].

**Treatment for T2DM: lessons from the TODAY study**

Despite the escalating rates of obesity-driven T2DM in youth, therapeutic options remain limited to metformin, the only FDA-approved oral hypoglycemic agent, and insulin when the former fails [44]. Even though metformin was effective in the short term over 16 weeks, it remained unknown whether this effect was durable [45] until the results of TODAY (Treatment Options for type 2 Diabetes in Adolescents and Youth) showed more than 50% failure rates on metformin [46]. The TODAY study was a multicenter, randomized (699 overweight youth between 10 and 17 years of age, with T2DM of

![Figure 55.2](image_url)
with obesity—a problem that is no longer restricted to minority groups. European Journal of Endocrinology 2004;151:199–206.


CHAPTER 56

Diabetes in pregnancy

Yoel Toledano, Eran Hadar and Moshe Hod
Division of Maternal Fetal Medicine, Helen Schneider Hospital for Women, Rabin Medical Center, Petah-Tiqva, Israel

Key points

• Pregnancy is characterized by insulin resistance, developing during the second half of gestation. When hyperinsulinemia cannot compensate for exaggerated postprandial hyperglycemia, gestational diabetes develops.

• The perturbations of glucose homeostasis in pregnancy are classified into two main categories: gestational (GDM) and pregestational diabetes mellitus.

• Generally, it is estimated that diabetes (type 1, type 2, and GDM) occurs annually in 1–14% of all pregnancies, and specifically in 6–7% of pregnancies in the United States.

• Selective or universal screening for gestational diabetes mellitus should be determined by population characteristics. A new method for diagnosis of gestational diabetes mellitus has been recently suggested.

• Pregnancy complicated by diabetes is associated with abortions, major congenital malformations, macrosomia, and increased perinatal mortality.

• The key to improving pregnancy outcomes in women with diabetes is strict glycemic control.

• Alternatives for gestational diabetes treatment are medical nutrition and physical activity therapy either alone or in combination with glyburide, or insulin.

• Continuous glucose monitoring technology uncovers previously undetected glucose perturbations and glucose trends and is a helpful adjunct to conventional therapy.

• Preconceptional evaluation for women with pregestational diabetes should be encouraged.

• Postpartum screening for type 2 diabetes (T2DM) should be performed at 6–8 weeks in all women with gestational diabetes mellitus.

Maternal adaptation to pregnancy

Variable physiologic changes begin shortly after implantation [1]. Hyperplasia and hypertrophy of the β cells in the islets of Langerhans are ascribed to stimulation by estrogen and progesterone. Glucose crosses the placenta by facilitated diffusion during early pregnancy to supply glucose requirements of the fetus. Maternal fasting hypoglycemia may occur as a result of this enhanced glucose transport. In contrast to a seemingly normal basal insulin secretion, postprandial insulin levels are elevated, mainly due to an increase in synthesis and secretion.

Insulin resistance, developing during the second half of pregnancy, is related to an elevation of several hormones, including human placental lactogen, glucocorticoids and progesterone, as well as free fatty acids and tumor necrosis factor-α. Furthermore, glucose ingestion in late pregnancy results in higher and more sustained levels of glucose and insulin compared to nonpregnant subjects. In addition, glucagon is suppressed to a greater degree.

Recently, the glycemic profile in the second half of nondiabetic pregnancy was characterized by studies using continuous glucose monitoring (CGM) [2]. Hence, it is currently possible to define better glycemic targets in order to mirror normoglycemia in pregnant diabetic women. In nonobese, non-diabetic, subjects, the difference in glucose levels between fasting and preprandial values during the day was minimal. This suggested that fasting plasma levels may reflect preprandial values [2]. However, mean blood glucose levels during the night-time (23:00–06:00) were significantly lower than daytime levels. In a recent study, CGM confirmed that diurnal glucose patterns throughout the day are reduced by 20% and also that the percent of time in hypoglycemia is significantly higher in the pregnant versus nonpregnant states [3].

Classification of glucose abnormalities in pregnancy

The perturbations of glucose homeostasis in pregnancy are classified into two main categories:

1. Gestational diabetes (GDM), first defined by O’Sullivan in 1961, as “carbohydrate intolerance of varying severity with onset or first recognition during pregnancy.” This condition,
Pregestational diabetes includes all women whose glucose intolerance is discovered before conception, including type 1 (T1DM) and type 2 diabetes mellitus (T2DM), without differentiating between them. They are further categorized according to length of disease and the presence of microvascular or other end-organ complications. Additionally, the ongoing epidemic of obesity and diabetes has led to increased incidence of T2DM in women of childbearing age. Hence, it is diagnosed more frequently during pregnancy in the last years. Consequently, it is reasonable to screen women with risk factors for T2DM already at their first prenatal visit, using standard criteria for the diagnosis of diabetes [4].

Epidemiology of diabetes in pregnancy

Based on both national birth statistics and small community studies, it is estimated that diabetes (T1, T2, and GDM) occurs annually in 1–14% of all pregnancies. Specifically, it has been demonstrated that approximately 6–7% of pregnancies in the United States are complicated by diabetes mellitus [5]. Actually, fetal or maternal complications are reported in an estimated 400,000 live births in the USA each year. More than 90% of them are linked to maternal glucose abnormalities, especially large-for-gestational-age infants and macrosomia [6]. When other complications are considered, including spontaneous abortion, fetal malformation (principally found in pregestational diabetes), shoulder dystocia, preeclampsia, and delivery by cesarean section, the incidence of diabetes in pregnancy rises to 13%. Moreover, GDM is a major public health problem with a growing magnitude. It is estimated at up to 20% by evidence collected from high-prevalence groups, such as ethnic minorities (American Indians, Native Hawaiians and other Pacific Islanders, and Hispanics).

It is estimated that between 0.4% and 2% of all births are complicated by pregestational diabetes. Indeed, differentiating the number of births complicated by T1DM from those complicated by T2DM is difficult as the prevalence of each of them varies by both race and age. Notably, it is assumed that the prevalence of pregnancies complicated by pregestational diabetes will increase due to several factors, including increased incidence of obesity in women of childbearing age, increased incidence in T2DM among adolescents and women younger than 25 years of age, increased birth rate among adolescents, and population growth among groups with higher risk of diabetes.

The prevalence of GDM is directly related to the prevalence of T2DM in a given population. GDM prevalence in the US is around 5–6%, accounting for at least 85% of diabetes occurring in pregnancy [5]. However, controversy in the epidemiology of GDM concerns debates on who should be screened, when screening should occur, screening test criteria, and diagnostic test criteria [6].

Table 56.1 presents risk factors for GDM. Accumulation of risk factors increases the risk. Moreover, the offspring of diabetic mothers are at an increased risk for future insulin resistance. Taking into account the increased prevalence of obesity and T2DM at this era, GDM occurrence may significantly rise, with a vicious cycle of insulin resistance, obesity, diabetes, and their consequent complications in future generations.

Diagnosis of gestational diabetes mellitus

The basis for the diagnosis of gestational diabetes mellitus (GDM) was laid down by O’Sullivan et al. in the 1960s [7]. They suggested the criteria for an abnormal 100 g oral glucose tolerance test (OGTT), based on an outcome set to identify women at risk for subsequent diabetes. Thus, the major shortcoming of those criteria is the lack of correlation to adverse pregnancy outcomes. The currently accepted guidelines for the diagnosis of GDM were presented in the Fourth International Workshop-Conference on GDM [8] and were endorsed by several key organizations [4,9]. Strategies for screening and diagnosis of GDM, will be discussed in the following sections, and are summarized in Figure 56.1 and Table 56.2.

Screening and its cost-effectiveness

Screening recommendations range from the inclusion of all pregnant women (universal) to the exclusion of all women except those at risk (selective). Screening policy should be determined locally, depending upon population characteristics, local prevalence and preference. It is achieved by either selecting high-risk populations or a 50 g glucose challenge test (GCT). The American Congress of Obstetricians and Gynecologists (ACOG) [9] recommends universal screening for all pregnant women, to be done between 24–28 weeks of gestation, except women who meet all the low-risk criteria. The American Diabetes Association (ADA) [4] supports selective screening depending upon risk factors (Table 56.1). Risk assessment

<table>
<thead>
<tr>
<th>Table 56.1 Risk factors for gestational diabetes mellitus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advanced maternal age</td>
</tr>
<tr>
<td>Maternal obesity</td>
</tr>
<tr>
<td>High parity</td>
</tr>
<tr>
<td>Previous delivery of a macrosomic infant</td>
</tr>
<tr>
<td>Family history of type 2 diabetes mellitus</td>
</tr>
<tr>
<td>Maternal short stature</td>
</tr>
<tr>
<td>Polycystic ovary disease</td>
</tr>
<tr>
<td>High levels of saturated fat in the diet</td>
</tr>
<tr>
<td>Prior gestational diabetes</td>
</tr>
<tr>
<td>Prior neonatal death</td>
</tr>
<tr>
<td>Prior cesarean delivery</td>
</tr>
<tr>
<td>Previous stillbirth or congenital malformations</td>
</tr>
<tr>
<td>High blood pressure during pregnancy</td>
</tr>
<tr>
<td>Multiple pregnancies</td>
</tr>
</tbody>
</table>

Screening recommendations range from the inclusion of all pregnant women (universal) to the exclusion of all women except those at risk (selective). Screening policy should be determined locally, depending upon population characteristics, local prevalence and preference. It is achieved by either selecting high-risk populations or a 50 g glucose challenge test (GCT). The American Congress of Obstetricians and Gynecologists (ACOG) [9] recommends universal screening for all pregnant women, to be done between 24–28 weeks of gestation, except women who meet all the low-risk criteria. The American Diabetes Association (ADA) [4] supports selective screening depending upon risk factors (Table 56.1). Risk assessment...
Diabetes in pregnancy

GDM risk assessment should be ascertained at the first prenatal visit, according to risk.

**Low risk criteria:**
- Member of an ethnic group with a low prevalence of GDM
- No known diabetes in first-degree relatives
- Age ≤ 25 years
- Normal weight before pregnancy
- Normal weight at birth
- No history of abnormal glucose metabolism
- No history of poor obstetric outcome

**High risk criteria:**
- Severe obesity
- Strong family history of type 2 diabetes
- Previous history of:
  - GDM, impaired glucose metabolism or glucosuria

If all low risk characteristics are present

If none of the low risk or high risk criteria is fulfilled

If any of the high risk characteristics is present

Blood glucose testing not routinely required

Blood glucose testing at 24–28 weeks

Blood glucose testing as soon as feasible

Two-step procedure: 50 g GCT followed, according to threshold, by an OGTT

One-step procedure: diagnostic OGTT performed on all subjects

If GDM is not diagnosed, blood glucose testing should be repeated at 24–28 weeks or at any time a patient has symptoms or signs suggestive of hyperglycemia

---

**Table 56.2** Diagnosis of gestational diabetes mellitus (GDM) by an oral glucose tolerance test (OGTT) [8]

<table>
<thead>
<tr>
<th></th>
<th>100 g glucose</th>
<th>75 g glucose</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Carpenter &amp; Coustan criteria(^a)</td>
<td>NDDG criteria(^b)</td>
</tr>
<tr>
<td></td>
<td>mg dL(^{-1})</td>
<td>mmol L(^{-1})</td>
</tr>
<tr>
<td>Fasting</td>
<td>95</td>
<td>5.3</td>
</tr>
<tr>
<td>1 hour</td>
<td>180</td>
<td>10.0</td>
</tr>
<tr>
<td>2 hour</td>
<td>155</td>
<td>8.6</td>
</tr>
<tr>
<td>3 hour</td>
<td>140</td>
<td>7.8</td>
</tr>
</tbody>
</table>

\(^a\) Plasma/serum glucose  
\(^b\) Plasma glucose
should be carried for all pregnant women at the first prenatal follow-up, to be made as early as possible during pregnancy (Figure 56.1). In contrast, other organizations have stated that there is insufficient evidence to make a recommendation for or against GDM screening [10].

The main screening tool is the GCT, preformed with or without overnight fasting, and involves plasma glucose measurement 1 hour after ingestion of 50 g glucose. Glucose concentrations of 130–140 mg/dL (7.2–7.8 mmol/L) are considered as the positive cutoff, indicating the need for a diagnostic OGTT. A threshold set to 130 mg/dL has a sensitivity of nearly 100% with 23% of women requiring an OGTT. A 140 mg/dL cutoff yields a sensitivity of nearly 90%, with a 15% positive screen rate [11]. Importantly, about 10% of those with screening results of 130–139 mg/dL will have GDM; however, the lower threshold is associated with a 12% increase in the cost for GDM diagnosis [12]. Because the precise cost–benefit ratio for diagnosing GDM remains unresolved, either threshold is acceptable. The selection should be decided upon local considerations. Thus, in high-prevalence countries it is reasonable to use the lower threshold (for increased sensitivity), and the higher cut-off may be favored (for lower cost) when the prevalence is low.

Recently, the cost-effectiveness of GDM screening and intervention was studied in India and Israel, settings with contrasting epidemiologic and cost environments [13]. Significantly, GDM interventions are highly cost-effective in both settings, by World Health Organization standards.

**Diagnosis of GDM: the traditional criteria**

The most commonly accepted method of GDM diagnosis is a 100 g OGTT, using either the National Diabetes Data Group (NDDG) criteria [14] or the Carpenter and Coustan criteria [15] (Table 56.2). The test should be performed in the morning after an 8–14-hour overnight fast and after at least 3 days of unrestricted diet (>150 g carbohydrate per day) and physical activity. Plasma glucose is measured at fasting, 1, 2, and 3 hours following glucose load. Two or more of glucose concentrations must be met or exceeded for a positive diagnosis. Alternatively, new criteria are listed in the following sections.

**The HAPO Study**

As previously mentioned, the diagnostic value of the O’Sullivan criteria and the derived guidelines are of limited value, as the true levels of hyperglycemia that have an impact on pregnancy outcomes, were not established. In order to establish an evidence base to the diagnosis of GDM, the hyperglycemia and adverse pregnancy outcome (HAPO) study was designed [16].

The HAPO was a prospective, observational, study held in 15 medical centers from 9 countries with 23,316 women. Each participant, at approximately 28 weeks, preformed a 2-h 75-g OGTT. Caregivers and participants remained blinded to the results unless glucose values indicated overt diabetes or hypoglycemia. The results demonstrated an association between increasing levels of fasting, 1-h and 2-h plasma glucose post a 75-g OGTT, to the four primary endpoints of the study: birth weight above the 90th percentile, cord blood C-peptide level above the 90th percentile, primary cesarean delivery, and clinical neonatal hypoglycemia. Positive correlations were also found between increasing plasma glucose levels to the five secondary outcomes: premature delivery, shoulder dystocia or birth injury, intensive neonatal care admission (NICU), hyperbilirubinemia, and preeclampsia.

**IADPSG criteria: a new approach for the diagnosis of diabetes in pregnancy**

Based on the results of the HAPO, with supplementary and ancillary investigations, the International Association of Diabetes and Pregnancy Study Groups (IADPSG) suggested a new strategy for diabetes diagnosis during pregnancy (Tables 56.3, 56.4, and 56.5) [4,17]. At the first prenatal visit, all (universal) or only high-risk women (selective) should be tested for either fasting plasma glucose (FPG), hemoglobin A1c (HbA1c) or random plasma glucose (RPG). High-risk criteria include: pre-pregnancy obesity, family history of T2DM, prior GDM and known carbohydrate intolerance or glucosuria. Recently, the value of FPG for the diagnosis of GDM at the first prenatal visit was evaluated in a Chinese population [18]. It was suggested that FPG value of 6.10–7.00 mmol L⁻¹ (110–126 mg dL⁻¹), and not 5.10–6.00 mmol L⁻¹ (92–109 mg dL⁻¹), at the first prenatal visit should be considered as the criterion for diagnosis of GDM.

The next phase, at 24–28 weeks of gestation, includes a single diagnostic step with a 2-h 75-g OGTT. The IADPSG criteria allow the diagnosis of overt diabetes to be made during pregnancy, an option not available in the previously mentioned guidelines. Importantly, the suggested cut-offs are lower than those currently acceptable, which will lead to a higher prevalence of GDM, but these cut-offs are based on evidence extrapolated from the HAPO study. At the present time, the debate as to whether or not to adopt the IADPSG criteria is still ongoing. The ADA has already adopted these recommendations [4]. The ACOG has issued a committee opinion, not supporting the IADPSG single-step recommendations [19] due to lack of evidence that the identification and treatment of women, based on the IADPSG criteria, will lead to clinically significant improvements in maternal and neonatal outcomes. Rather, this management can cause a significant increase in healthcare costs. The WHO and the National Institutes of Health (NIH) are expected to publish guidelines in the near future.

Using the Israeli HAPO study data, Kalter-Leibovici et al. [20] found that adoption of IADPSG recommendations would increase GDM diagnosis by ∼50% (from 6% to 9%). Risk stratification may reduce overtreatment in one third of IADPSG-positive women. FPG ≥ 89 mg dL⁻¹ or BMI ≥ 33.5 kg m⁻² at 28–32 weeks of gestation detected proportions of adverse outcomes similar to IADPSG criteria. Hence, FPG or BMI may be a practical screening alternative.
Table 56.3  Strategy for the detection and diagnosis of hyperglycemia disorder in pregnancy [17]

<table>
<thead>
<tr>
<th>1st prenatal visit</th>
<th>Measure fasting plasma glucose (FPG), hemoglobin A1c (HbA1c), or random plasma glucose (RPG) on all or only high-risk women.</th>
</tr>
</thead>
<tbody>
<tr>
<td>24–28 weeks of gestation</td>
<td>75 g OGTT on all women not previously diagnosed with overt diabetes or GDM</td>
</tr>
</tbody>
</table>

If results indicate overt diabetes as per Table 56.4
→ Preexisting (overt) diabetes

If results not diagnostic of overt diabetes as per Table 56.4 and FPG ≥ 7.0 mmol L⁻¹ (126 mg dL⁻¹)
→ GDM

If results not diagnostic of overt diabetes as per Table 56.4 and FPG < 5.1 mmol L⁻¹ (92 mg dL⁻¹)
→ test for GDM during 24–28 weeks, with a 75 g OGTT

If one or more values equals or exceeds thresholds as per Table 56.5
→ GDM

If all values less than thresholds indicated as per Table 56.5
→ Normal

Table 56.4  Threshold values for diagnosis of overt diabetes in pregnancy [17]

<table>
<thead>
<tr>
<th>Measure of glycemia</th>
<th>Threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fasting plasma glucose</td>
<td>≥ 7.0 mmol L⁻¹ (126 mg dL⁻¹)</td>
</tr>
<tr>
<td>Hemoglobin A1c</td>
<td>≥ 6.5%</td>
</tr>
<tr>
<td>Random plasma glucose</td>
<td>≥ 11.1 mmol L⁻¹ (200 mg dL⁻¹)</td>
</tr>
</tbody>
</table>

If a random plasma glucose is the initial measure of glycemia, the tentative diagnosis of overt (preexisting) diabetes should be confirmed by fasting plasma glucose or hemoglobin A1c

Table 56.5  Glucose threshold values for diagnosis of GDM [17]

<table>
<thead>
<tr>
<th>Glucose measure</th>
<th>Glucose threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fasting plasma glucose</td>
<td>5.1 mmol L⁻¹ (92 mg dL⁻¹)</td>
</tr>
<tr>
<td>1-hour post 75-g OGTT</td>
<td>10.0 mmol L⁻¹ (180 mg dL⁻¹)</td>
</tr>
<tr>
<td>2-hour post 75-g OGTT</td>
<td>8.5 mmol L⁻¹ (153 mg dL⁻¹)</td>
</tr>
</tbody>
</table>

Adverse pregnancy outcomes in gestations complicated by diabetes

Pregnancies complicated by diabetes, whether pregestational or gestational, are subject to short- and long-term complications which may arise throughout pregnancy, delivery, and postpartum. Adverse outcomes may inflict the diabetic mother as well as her fetus, neonate, child, and adult offspring (Table 56.6).

Glucose and possibly other metabolites account for the pathogenesis of these adverse outcomes, as they create an altered environment in which the fetus is exposed to changes in gene expression, free oxygen radicals, cellular damage, impaired metabolic environment and increased teratogenesis [21]. Hence, the rate of these complications is closely related to the degree of pregestational, periconceptional, and antenatal glycemic control.

Congenital anomalies

Major congenital anomalies are the primary contributor to perinatal mortality and morbidity for the offspring of a mother with pregestational diabetes, accounting for 30–50% of perinatal mortality. There is a three- to fivefold increase in the incidence of congenital anomalies in infants of pregestational diabetic mothers, rising from approximately 2% incidence in the normal population to 6–11%. The congenital malformation rate is inversely related to maternal age at onset of diabetes and directly related to the level of glycemic control prior to pregnancy and during early pregnancy [22]. Poor glycemic control exposes abnormal levels of glucose, insulin and ketones which are involved in the pathogenesis of malformations. Anomalies occur early in pregnancy, during the period of organogenesis, usually up until the 7th week of gestation. They do not involve a specific organ. Rather, cardiac, central nervous, genitourinary and gastrointestinal systems may be implicated. Specific anomalies include neural tube defects, holoprosencephaly, ventricular septal defects, and transposition of the great vessels. Caudal dysplasia is highly characteristic, and although occurs more frequently in diabetic pregnancies, it is still rare and not pathognomonic. The critical time for glycemic control is at a time period where women are either not pregnant, unaware of their pregnancy or are aware but still not seeking prenatal care. Therefore, management and
Table 56.6 Gestational and pregestational diabetes-associated morbidity during pregnancy

<table>
<thead>
<tr>
<th>Maternal</th>
<th>Fetal</th>
<th>Neonatal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypertensive disorders</td>
<td>Abortion</td>
<td>Short term (baby)</td>
</tr>
<tr>
<td>Urinary and genital tract infections</td>
<td>Stillbirth</td>
<td>Respiratory morbidity</td>
</tr>
<tr>
<td>Cesarean section</td>
<td>Congenital anomalies</td>
<td>Neonatal intensive care unit admission</td>
</tr>
<tr>
<td>Future type 2 diabetes mellitus</td>
<td>Large-for-gestational-age and macrosomia</td>
<td>Hypoglycemia</td>
</tr>
<tr>
<td></td>
<td>Preterm labor and delivery</td>
<td>Hyperbilirubinemia</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hypocalcemia</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polyhydramnios</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instrumental delivery</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Traumatic labor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Birth injury and shoulder dystocia</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Counseling of women with diabetes should begin before conception. Unfortunately, less than 20% of diabetic women undergo pre-pregnancy counseling. Lack of preconceptional care and pre-pregnancy uncontrolled diabetes are associated with an increased malformation rate. Appropriate pregnancy planning and glycemic control may decrease the rate of congenital anomalies to a comparable rate of the background population. Mean glucose level \(< 100 \text{ mg dL}^{-1}\) (5.6 mmol L\(^{-1}\)) is associated with a low incidence of anomalies, as well as HbA1c of less that 7%.

**Spontaneous abortions**
The risk for a spontaneous abortion is also correlated to the degree of glycemic control. Good control, with normal HbA1c, reduces the risk of miscarriage to similar levels of the non-diabetic population.

**Stillbirth**
Stillbirth was a common complication in diabetic pregnancies and has become a rare event in current practice, nonetheless devastating. Hyperglycemia is involved in the pathogenesis of several factors leading to stillbirth—mainly congenital anomalies, fetal hypoxemia and acidemia, placental insufficiency, and macrosomia. Maternal hyperglycemia is translated into hyperinsulinemia, mediated by hyperplasia of fetal pancreatic islet cells. Pedersen and others [23,24] have linked it to cord blood acidemia and hypoxemia, which are related to increased rates of stillbirth. As for other complications, stillbirth is inversely related to the level of glycemic control. Mean blood glucose \(<100 \text{ mg dL}^{-1}\) is associated with low perinatal mortality rate.

Fetal demise related to inappropriate metabolic control occurs during late 3rd trimester, and is more pronounced in pregnancies complicated with hypertension, vascular complications, polyhydramnios, growth restriction, and mainly in those with T1 and T2DM.

**Macrosomia and large-for-gestational-age**
Macrosomia is defined as birth weight above 4000 g and large-for-gestational-age (LGA) as birth weight above the 90th percentile, according to sex and gestational age. Macrosomia is related to increased risk of stillbirth, shoulder dystocia and birth trauma, and an increased rate of operative and cesarean deliveries. Good glycemic control may decrease the rates of macrosomia from 30–50% to as low as 0–22%, depending on glucose values. Tight glycemic control leads to 0%, 9% and 11% macrosomia rates with respective mean blood glucose values of 80–87 mg dL\(^{-1}\), less than 110 mg dL\(^{-1}\), and 105–121 mg dL\(^{-1}\) [25–27]. It has been suggested that 3rd trimester nonfasting glucose values, especially the 1-h postprandial glucose is best correlated with the risk of macrosomia.

**Detection and treatment of diabetes in pregnancy**
Diabetes in pregnancy is a special situation. On one hand, it is important to control hyperglycemia in the mother. On the other hand, preventing the deleterious effects of sugar excursions on the developing fetus is crucial. Until 1980 physicians were advising diabetic women against childbearing due to grave consequences of diabetes in pregnancy. However, as medical knowledge and technologies for treating diabetes before, during, and after pregnancies improved, the prognosis of such pregnancies has greatly improved.

**Glycemic control targets**
The single most important step to achieving minimal morbidity and mortality rates in GDM is to establish near-normal metabolic control. The complication most clearly correlated to maternal hyperglycemia, specifically maternal postprandial hyperglycemia, is fetal macrosomia. In fact, controlling postprandial glucose levels reduces macrosomia, neonatal hypoglycemia, and cesarean delivery as compared to managing preprandial glucose levels alone [28]. From these studies it is apparent that therapy for pregnant women with diabetes should be targeted to treat 1–2-h postprandial glucose coincident with
Diabetes in pregnancy

Table 56.7 Glucose target values and regular insulin dose during labor and delivery

<table>
<thead>
<tr>
<th>Insulin dose (units)</th>
<th>Glucose levels</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mg dL(^{-1})</td>
<td>mmol L(^{-1})</td>
</tr>
<tr>
<td>1–2 units per hour</td>
<td>95–150</td>
<td>5.3–8.3</td>
</tr>
<tr>
<td>2 units i.v. push</td>
<td>151–180</td>
<td>8.4–10.0</td>
</tr>
<tr>
<td>3 units i.v. push</td>
<td>181–220</td>
<td>10.1–12.2</td>
</tr>
<tr>
<td>4 units i.v. push</td>
<td>&gt;220</td>
<td>&gt;12.2</td>
</tr>
</tbody>
</table>

the postprandial hyperglycemia peak (Table 56.8). Nevertheless, controlling of preprandial and nocturnal glucose levels is of paramount importance as well (Table 56.8).

Preconception care for pregestational diabetes

Preconception counselling for women with pregestational diabetes mellitus is of paramount importance and should be encouraged. As congenital anomalies and spontaneous abortions are serious complications in pregestational diabetes, a preconception counselling is beneficial and cost-effective. If strict glycemic control can be safely achieved, the recommended pre-pregnancy HbA1c targets are <6.1% (NICE) [29] and <7% (ADA) [4]. An evaluation of glycemic control, hypoglycemic symptoms and blood pressure should be performed. A survey for diabetes vascular complications is advisable, including a retinal and neurologic examination, a renal function (eGFR or creatinine clearance), a urinary microalbumin excretion test, and, in selected patients, an electrocardiogram. Thyroid function studies should be obtained in patients with T1DM due to comorbidity. Folic acid should always be prescribed as part of preconception care, especially in women with diabetes given their increased risk of neural tube defects.

Diet and behavioral adjustments

Diet is the foundation of treatment for all types of diabetes. Importantly, patient education and adapted physical exercise routines should be combined with the diet plan. Currently, two nutritional approaches are recommended: decreasing the proportion of carbohydrates to 35–40% in a daily regimen of three meals and three or four snacks, and lowering glycemic index carbohydrates for approximately 60% of daily intake. A regimen of smaller, more frequent, meals leads to better satiety and compliance, with a reduction of postprandial glucose peaks. Of note, insulin resistance is highest in the morning. Postprandial glucose levels are directly influenced by the amount of carbohydrates in the consumed food. Therefore, carbohydrate-based calories should be consumed at later times of the day, and breakfast should only be a small meal.

The assignment of daily caloric intake is similar for gestational and pregestational diabetes. It is calculated based on pre-pregnancy BMI and according to actual pregnancy weight (20–25 kcal kg\(^{-1}\) for obese women and 30–35 kcal kg\(^{-1}\) for nonobese women). Postprandial blood glucose levels and insulin sensitivity may improve control by an appropriate exercise program. However, pregnant diabetic women need to be willing as well as able (socioeconomic limitations, obesity, multiparity) to participate. Pharmacologic agents should be prescribed for patients with GDM who fail to achieve the desired level of glycemic control.

Table 56.8 Capillary blood glucose and HbA1c target values for patients with diabetes in pregnancy

<table>
<thead>
<tr>
<th></th>
<th>ACOG recommendations</th>
<th>ADA recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GDM</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fasting glucose</td>
<td>≤95 (5.3)</td>
<td>≤95 (5.3)</td>
</tr>
<tr>
<td>1-hour postprandial glucose</td>
<td>≤130–140 (7.2–7.8)</td>
<td>≤140 (7.8)</td>
</tr>
<tr>
<td>2-hour postprandial glucose</td>
<td>≤120 (6.7)</td>
<td>≤120 (6.7)</td>
</tr>
<tr>
<td><strong>Pregestational diabetes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HbA1c, %</td>
<td>≤6</td>
<td>&lt;6</td>
</tr>
<tr>
<td>Capillary blood glucose, mg dL(^{-1}) (mmol L(^{-1}))</td>
<td>≤95 (5.3)</td>
<td>60–99 (3.3–5.5)</td>
</tr>
<tr>
<td>Fasting glucose</td>
<td>≤100 (5.6)</td>
<td>60–99 (3.3–5.5)</td>
</tr>
<tr>
<td>Preprandial glucose</td>
<td>≤140 (7.8)</td>
<td>100–129 (5.6–7.2)</td>
</tr>
<tr>
<td>1-hour postprandial glucose</td>
<td>≤120 (6.7)</td>
<td>100–129 (5.6–7.2)</td>
</tr>
<tr>
<td>2-hour postprandial glucose</td>
<td>≤100 (5.6)</td>
<td></td>
</tr>
<tr>
<td>Mean capillary glucose</td>
<td>≤100 (5.6)</td>
<td></td>
</tr>
<tr>
<td>Bedtime glucose</td>
<td>Not &lt;60 (3.3)</td>
<td>60–99 (3.3–5.5)</td>
</tr>
<tr>
<td>Overnight glucose</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Insulin treatment**

There is a gradual increase in insulin requirement throughout pregnancy: 0.7 units kg\(^{-1}\) d\(^{-1}\) in the first trimester; 0.8 units kg\(^{-1}\) d\(^{-1}\) at week 18, 0.9 units kg\(^{-1}\) d\(^{-1}\) at week 26, and 1.0 units kg\(^{-1}\) d\(^{-1}\) from week 36 until delivery. In order to mimic the physiologic insulin secretion throughout the day, it is essential to pair rapid-acting insulin analogues with intermediate or long-acting insulins.

**Long-acting insulin analogues**

Two insulin analogues have a prolonged basal activity: insulins detemir and glargine. They function as basal, non-meal-related, insulin during the day. Only detemir was studied before and during pregnancy in a large randomized controlled trial of patients with T1DM [30]. Treatment with detemir resulted in a noninferior HbA1c during late gestation and similar maternal hypoglycemia rates compared with NPH insulin. In addition, perinatal morbidity and mortality were comparable, and no specific safety issues were identified [31]. However, several retrospective analyses have demonstrated that glargine may also be safe in pregnancy [32–34]. Indeed, results from a transplacental transfer study showed that insulin glargine has not crossed the placenta at therapeutic concentrations [35]. Degludec is a new ultralong-acting insulin on the horizon. Randomized controlled studies should be established for its use in pregnancy.

**Intermediate-acting insulin**

Neutral protamine Hagedorn (NPH) insulin has an 8-hour duration. Thus, it is an intermediate-acting insulin. A basal insulin action in a pregnant woman can be attained by three injections of NPH. However, the risk of hypoglycemia is increased, particularly at night, due to an insulin peak in the NPH profile. Currently, the best basal insulin action can be achieved by using insulin detemir or insulin aspart/lispro as a constant infusion in an insulin pump.

**Short-acting insulin**

Regular human insulin is inferior to the rapid-acting insulin analogues as its action is too slow to control the peak postprandial glucose and the risk of late postprandial hypoglycemia is increased. This mismatch between insulin to meal can be ameliorated by prescribing a very low carbohydrate meal and a large, between-meal, snack.

**Rapid-acting insulin analogues**

The rapid-acting insulins lispro and aspart have greater postprandial control, reduced hypoglycemia and greater convenience of administration than regular insulin. Insulin glulisine has not been studied in pregnancy.

*Insulin lispro* has a rapid onset (10–30 min) and a shorter duration of action (3–5 h) in comparison to human insulin. Studies have suggested several advantages of lispro compared to regular insulin: (a) improves postprandial glycemia in women with GDM [36]; (b) neonates of mothers on lispro displayed lower cranial/thoracic circumferences averaging 10th to 25th percentiles [36]; (c) lowers predelivery HbA1c in women with GDM [37]; (d) unlikely to cross the placenta [38].

In contrast, other reports suggested congenital anomalies with the use of lispro [39]. However, this association is probably due to poor glycemic control rather than drug-induced teratogenesis [40].

*Insulin aspart* also has faster absorption compared to regular human insulin. A study of aspart versus human insulin in T1 diabetic pregnant women demonstrated greater postprandial control, less nocturnal hypoglycemia, and reduced fetal loss and preterm deliveries [41,42].

**Safety of insulin analogues**

A concern for the use of insulin and its analogues was raised regarding their mitogenic effects. Insulin-like growth factor-1 receptor (IGF-1R) is a protein structurally similar to the insulin receptor. The individual level of IGF-1R of each cell was found to have an important role in the ability of insulin to promote growth. Insulin aspart and detemir exhibited 18% and 84% decrease in IGF-1 affinity, respectively, compared to human insulin [43]. In contrast, lispro and glargine had elevated affinities of 1.5- and 6.5-fold, respectively, compared to human insulin. Thus, mitogenic potential of glargine should be further studied.

**Oral antidiabetic agents**

*Glyburide* transfers minimally the placenta. Even after it crosses the placenta, glyburide may be actively effluxed from the fetus to the mother [44]. In studies by Langer et al., no association has been established between maternal glyburide use and macrosomia, neonatal hypoglycemia, other neonatal morbidity, or cord serum insulin concentration [45,46].

*Metformin* has been shown to cross the placenta [47]. It was used in several states: (a) To increase fertility and reduce spontaneous abortions in nondiabetic women with polycystic ovarian syndrome. In these women, metformin was not associated with teratogenesis and adverse growth, motor and social development at 3- and 6-month follow-ups [48,49]. (b) To control non-insulin-dependent diabetes in the first trimester. It has been shown to be relatively safe and with no increased risk in congenital anomalies. (c) To control GDM during the second and third trimester of GDM. No adverse perinatal outcomes were observed when compared to insulin treatment [50,51].

Overall, the complete safety and efficacy of oral antidiabetic agents in pregnancy cannot yet be conclusive. Metformin has less utility as a prospective monotherapy than glyburide since fewer women on the latter therapy require supplement insulin to maintain glycemic control. Taken together, insulin therapy with intermediate/long and rapid-acting profiles is strongly recommended.
for the treatment of diabetes during pregnancy. When pharma-
ocologic treatment is required in patients with pregestational dia-
abetes, insulin should be used exclusively.

**New technologies**

**Continuous subcutaneous insulin infusion (CSII)
(Insulin pump)**

CSII is an alternative treatment to multiple daily injections 
(MDI) for people with T1DM. The new and modern pumps 
are smaller, more efficacious, easier to use and safe. However, 
no clear difference between CSII and MDI has been established 
for treatment in pregnant women with T1DM. Rather, both 
treatments are safe and valid during pregnancy [52]. Neverthe-
less, CSII treatment improves patients’ satisfaction and lifestyle 
flexibility compared to MDI.

**Continuous glucose monitoring system (CGMS)**

As previously mentioned, CGMS use during the second half of 
pregnancy enabled the definition of normal glycemia. Next, the 
time interval from meal to peak postprandial glucose levels was 
evaluated. Whereas in nondiabetic pregnancy the interval was 
70 min, it was somewhat later, ~90 min, in gestations compli-
cated by diabetes [2,53]. In addition, the peak glucose values 
were $103 \pm 26 \text{ mg dL}^{-1}$ and $164 \pm 53 \text{ mg dL}^{-1}$ in patients with 
well-controlled and poor-controlled diabetes, respectively. Tak-
ing into account the glycemic targets currently recommended 
for diabetes in pregnancy ($140 \text{ mg dL}^{-1}$ at 1 h and $120 \text{ mg dL}^{-1}$ at 
2 h, Table 56.8), it was suggested by Chitayat et al. [40] that blood 
glucose determinations should be taken at 90 min postmeal with 
a desired glucose value of $\sim 110 \text{ mg dL}^{-1}$. Importantly, this sug-
gested glucose target should be further tested for its correlation 
to pregnancy outcomes.

CGM was studied for several aspects of diabetes treatment 
in pregnancy. First, using CGM for treatment assessment, it 
was shown that undetected hyperglycemia and asymptomatic 
hypoglycemic events are common during pharmacologic 
treatment in GDM [54,55]. Second, “blinded” CGM for 72 hours 
was used for adjustment of therapy in pregnant women with 
T1DM [56]. It was concluded that this CGM method is not 
advisable as there is a wide variability in the day-to-day 
glucose levels of T1DM pregnancies. Hence, real-time CGM 
use could possibly overcome this limitation. Consequently, a 
study was designed with a sample size large enough to provide 
data on pregnancy outcomes in T1DM using real-time CGM. 
This CONCEPT study in pregnant women or women planning 
pregnancy has already been launched in 2013.

Future technological development consists of an artificial 
pancreas, based on a “closed-loop” system, i.e. delivering 
insulin and/or glucagon/glucose, sensing glucose and control-
ling delivery according to blood glucose levels [40]. However, 
the accuracy of the current CGM and algorithms is not yet 
sufficient to permit the loop to be closed. Thus, research 
direction should include the improvement of CGM and insulin 
algorithms. Implementation of an artificial pancreas for diabetes 
in pregnancy will hopefully lead to a perinatal outcome similar 
to nondiabetic subjects.

**Special considerations during pregnancy, 
labor, and delivery**

**Fetal assessment**

There are some important considerations for fetal assessment, 
during prenatal follow-up:

1. **Determination of gestational age**: As for any pregnancy, 
especially those complicated by diabetes, women should 
undergo a sonographic evaluation during early first trimester. 
The purpose is confirmation of viability, due to the higher 
risk of fetal loss, and accurate validation or determination 
of gestational age, because of possible emergence of fetal 
growth abnormalities later in pregnancy, and the goal of term 
delivery [57].

2. **Detection of congenital anomalies**: It is of importance and 
should be initiated as early as the first trimester and repeated 
in the second trimester of pregnancy. There is a considerable 
controversy regarding the optimal time at which to perform the 
sonographic examination aimed to detect structural anoma-
lies. The NICE antenatal care guidelines recommend that ultrasound screening for fetal anomalies should be routinely 
offered, between 18 to 21 weeks [57]. Others recommend 
performing two anomaly scans: the first at 10–14 weeks of 
gestation and the second at 20–22 weeks.

3. **Fetal well-being**: There are two main tests used in mod-
erm obstetrics to assess and monitor fetal well-being: fetal 
cardiocotography (using the nonstress test, NST) and the 
biophysical profile (BPP). The application of these tests and the 
reassurance they provide may prevent unnecessary early 
interventions and allow prolongation of pregnancy to term. 
The ACOG recommends that antenatal testing be performed 
on all those with poor glycemic control or any other related 
pregnancy complications [9]. The method of testing (NST 
or BPP) is left to the discretion of the provider guided by 
local practice. There is no unified opinion regarding the 
need for antepartum fetal assessment in women who have 
well-controlled, uncomplicated, preexisting or gestational 
diabetes. There is also no consensus on the appropriate timing 
to initiate testing, and the intervals to do so. It is usually 
started at 32–34 weeks on a weekly basis, but testing at earlier 
gestational ages and in shorter intervals, may be warranted 
in some pregnancies complicated by additional high-risk 
conditions.

4. **Surveillance of fetal growth**: It is important to detect cases 
in which fetal growth is deviant: either small or large for 
gestational age. Estimated fetal weight (EFW) employing vari-
ous formulas is the most commonly used method. Although the 
accuracy of fetal weight estimation decreases with increasing
birth weight there is no significant difference in the absolute error of birth weight in infants born to diabetic mothers and those born to women without diabetes, at about 10%.

Abdominal circumference (AC) is another method to estimate fetal growth, for the evaluation of metabolic control as well as the risk for macrosomia and shoulder dystocia. No difference was found between accuracy of EFW and that of abdominal circumference [58]. AC measurements combined with serial measurement of maternal blood glucose levels are a good indication of metabolic control and help modify treatment. Women with fetal AC > 75th percentile should be offered a tighter regime for glycemic control. Another form of deviant fetal growth is fetal growth restriction, usually defined as birth weight lower than the 10th percentile for a given gestational age and sex. Growth restriction is not a common complication of diabetic pregnancies and is more likely to occur in pregnancies of diabetic women suffering from severe vasculopathy.

**Timing of delivery**

The decision to deliver a patient with diabetes depends on a balance between potential complications of stillbirth, macrosomia, and shoulder dystocia against the risk of premature delivery as well as potential complications of labor induction. A summary of a decision analysis for timing and mode of delivery in pregnancies complicated by diabetes is presented in Figure 56.2. Maternal metabolic control, gestational age, and fetal growth assessment are all factors that should be considered. As a general rule, labor and vaginal delivery are aimed to occur spontaneously at term, after fetal maturation is achieved, provided that the patient’s diabetes is well controlled and antepartum surveillance is normal. Fetal indications favoring a planned delivery include macrosomia, LGA, previous stillbirth as well as prevention of stillbirth and reduction in shoulder dystocia. Maternal indications for a planned delivery include hypertension, diabetic vasculopathy, and poor compliance with the diabetic management resulting in adverse glycemic control.

In patients who fail to meet targeted glycemic control levels, normal fetal surveillance tests and no other immediate/emergent indication for delivery, a positive lung maturation test should precede the decision to deliver. If test results are negative, patients should be managed expectantly with fetal surveillance tests and lung maturation test repeated in a week. When diabetes is well controlled and gestational age is well documented, RDS at or beyond 39 weeks of gestation is rare enough that routine amniocentesis for pulmonary maturity is not necessary.

**Mode of delivery**

The optimal mode of delivery for a diabetic mother remains undecided. High rates of shoulder dystocia and Erb's palsy have prompted adoption of early induction to achieve vaginal delivery or a primary cesarean section in selected patients, based on estimated fetal size. Such approaches are limited by the relative inaccuracy of ultrasound prediction of birth weight, and by the high number needed to treat in order to prevent a single adverse outcome.

Cesarean delivery to prevent traumatic birth injury may be considered, as avoiding vaginal delivery when macrosomia is suspected reduces the rate of shoulder dystocia, Erb's palsy, or brachial plexus injury. Rouse et al. [59] calculated the probability of shoulder dystocia based on birth weight in diabetic and nondiabetic pregnancies. For birth weights of 4500 g or more there was a 52% probability in diabetic compared to 14% in nondiabetic pregnancies, and the mean probability that a neonatal brachial plexus injury would persist was 6.7 % (range 0−19%). Thus, to prevent one case of permanent brachial plexus injury in babies weighing 4500 g or more would necessitate 153 cesarean deliveries in diabetic mothers and 419 in nondiabetic mothers. If a cut-off of 4000 g is used, then 169 cesarean sections would be required in diabetic women compared to 654 in nondiabetic women. Conway et al. [60] conducted elective cesarean delivery in EFW > 4250 g and offered induction of labor if EFW was lower. Shoulder dystocia was reduced from 2.8% to 1.5% (OR 0.5 range 0.3–1.0). The cesarean delivery rate has increased from 21.7% to 25.1%. The ACOG recommends that an EFW > 4500 g is an indication for cesarean delivery.

The Cochrane database [61] compared a policy of elective induction at 38 weeks to expectant management up to 42 weeks. There was no difference in cesarean delivery rates between the groups (RR = 0.81; 95% CI 0.52−1.26), the induction group had a lower risk of macrosomia (RR = 0.56; 95% CI 0.32−0.98) and there were three cases of mild shoulder dystocia in the expectant management group. A large retrospective cohort study [62] conducted in the USA encompassing data for more than 100,000 deliveries found that although women with diabetes had an increased risk of cesarean delivery compared to women without diabetes (OR 2.00; 95% CI 1.83−2.19), induction of labor was associated with a lower risk of cesarean delivery.
compared to those whose labor was not induced (OR = 0.77; 95% CI 0.5–0.89). Kjos et al. demonstrated that induction at 38 weeks in a population of women with GDM was associated with a lower frequency of LGA and shoulder dystocia without an increased rate of cesarean delivery [63]. This is in contrast to studies of induction in nondiabetic women in which suspected macrosomia is apparently associated with an increased rate of cesarean delivery.

**Glycemic control during labor**

Neonatal hypoglycemia develops due to fetal hyperinsulinemia. Following birth, in the absence of continuous exposure to sufficient levels of glucose, the newborn may become hypoglycemic. Several trials have demonstrated that maternal hyperglycemia during labor and delivery is associated with neonatal hypoglycemia [64], birth asphyxia, and nonreassuring fetal heart rate tracings. Targeting maternal glucose levels to values of 4–7 mmol L⁻¹ (72–126 mg dL⁻¹) during labor, leads to a lower risk of maternal hypoglycemia than targeting lower levels.

**Postpartum care of women with diabetes**

Women with gestational diabetes mellitus, even after delivery, remain a high-risk group for the development of impaired fasting glucose, impaired glucose tolerance, diabetes mellitus (mostly type 2, but to a lesser extent even type 1), and the full spectrum of the metabolic syndrome. The onset of postpartum T2DM is observed within 5–10 years. The rate is reported from 2.6–70%, from 6 weeks to 28 years postpartum [65]. A multitude of studies reported various risk factors that may predict postpartum diabetes, including: ethnicity, fasting glucose values at OGTT and during pregnancy, postpartum fasting glucose, insulin use during pregnancy, maternal age, weight, BMI, HbA1c, previous history of gestational diabetes, family history of diabetes and parity [65,66]. Fasting glucose alone may not identify all women with glucose intolerance, and a formal glucose tolerance test is needed by a 75-g OGTT 6–12 weeks after delivery, and may be repeated periodically. Recurrence rates for gestational diabetes are 30–84%. Therefore, women with previous GDM, planning future pregnancies should be consulted appropriately prior to their following pregnancy.

**Breastfeeding**

All women with GDM or overt diabetes should be encouraged to breastfeed. Breastfeeding has been shown as a protective factor for the occurrence of fetal and maternal complications, and may reduce childhood obesity and T2DM. Also, breastfeeding may have maternal advantages, and postpartum lactation may be associated with lower glucose levels, weight loss, and lower rates of T2DM.

**Postpartum contraception**

Using combined oral contraception in T1DM does not influence glycemic control or promote end organ damage. Women with a history of gestational diabetes mellitus show no increase in the risk for T2DM, when taking the pill. Studies of the contraceptive patch and vaginal ring show that their metabolic effects are similar to those of the oral contraceptives, and they may be assumed safe for women with previous gestational diabetes. Copper IUD safety was demonstrated for healthy, T1 and T2DM women. The levonorgestrel-releasing IUD was studied in healthy and T1DM women and was not associated with metabolic disturbances. The data on the effect of progestin-only contraceptive is controversial, as some suggested an increased risk for T2DM in breastfeeding women with prior gestational diabetes, while others have not shown an effect on glucose values.

**References**

Maternal-Fetal and Neonatal Medicine 2013; Feb 14. [Epub ahead of print]


CHAPTER 57
Aging and diabetes mellitus

Hermes Florez¹, Jorge Ruiz² and Medha Munshi³

¹Divisions of Geriatric Medicine, Endocrinology & Epidemiology, Departments of Medicine & Public Health Sciences, University of Miami Miller School of Medicine, GRECC-Miami VA Healthcare System Miami, FL, USA
²Division of Geriatric Medicine, Department of Medicine, University of Miami Miller School of Medicine; Laboratory of E-learning and Multimedia Research, GRECC-Miami VA Healthcare System Miami, FL, USA
³Joslin Geriatric Diabetes Programs, Beth Israel Deaconess Medical Center, Department of Medicine, Harvard Medical School, Boston, MA, USA

Key points
- Aging is a major driver of the diabetes pandemic in addition to other risk factors, such as obesity and reduced physical activity.
- The growing impact of the diabetes epidemic in elderly populations affects both developed and developing countries.
- Diabetes in older adults is associated with higher mortality, reduced functional status, increased risk of institutionalization, as well as an enormous human and economic burden.
- Several conditions known as geriatric syndromes are common in older adults with diabetes and may have an impact on management and quality of life.
- The projected increase in diabetes prevalence in the elderly warrants more research and better healthcare planning to prevent complications, preserve functionality, and promote better quality of life.

Diabetes epidemiology in older adults

There are growing proportions of older adults among those with diabetes worldwide. Based on the 2012 International Diabetes Federation (IDF) Diabetes Atlas, it is estimated that 34.5% of the overall population with diabetes in 2012 were aged 60 – 79 years [1]. This proportion is expected to reach 41% by year 2030 and more than 252 million older adults with diabetes by 2035, according to estimates from the 2013 IDF Diabetes Atlas. The IDF regions and countries with the largest projected increase in the proportion of elderly patients with diabetes by 2030 are: Western Pacific (China and Indonesia), South & Central America (Colombia and Brazil), North America & Caribbean (Mexico, Canada, and USA), and Southeast Asia (Sri Lanka, Bangladesh, and India) (Table 57.1). Projections in the US suggest that the number of cases of diagnosed diabetes in those over 65 years of age will increase by 4.5-fold between 2005 and 2050 [2].

Data from the 2013 IDF Diabetes Atlas also showed near 135 million elderly patients with diabetes worldwide (with an overall prevalence of 18.6%). The countries with the largest number of older adults with diabetes were China (~33 million people, 20.6% prevalence), India (~15 million people, 16.2% prevalence), and the US (~10 million people, 21% prevalence), although data from the Centers for Disease Control and Prevention (CDC) has estimated than more than 25% of the population aged 65 years and older has diabetes in the US [3]. Population-based data in Western Europe showed approximately 25% diabetes prevalence in the age group 70 – 79 years [4]. However, the countries with the largest prevalence of diabetes in the elderly in 2012 were Egypt (32.9%), Mexico (29.6%), and Brazil (27.8%).

The prevalence of diabetes among elderly patients also varies depending on the diagnostic criterion used. Postprandial hyperglycemia, likely related to physiologic reductions in insulin secretion with aging [5,6], is a prominent characteristic of type 2 diabetes (T2DM) in older adults [7]. Using the A1c or FPG diagnostic criteria, as it is currently done for national surveillance, one third of older adults with diabetes are undiagnosed in the US [3]. Reports in Europe have supported the value of adding 2-hour glucose measurements (post glucose load) to identify incident diabetes in older populations [8].

Table 57.1 Top 10 countries with the greatest increase in number of older people with diabetes by 2030 [1]

<table>
<thead>
<tr>
<th>Country/territory</th>
<th>2012</th>
<th>% of elderly among adults with diabetes</th>
<th>2013</th>
<th>% of elderly among adults with diabetes</th>
<th>2030</th>
<th>% of elderly among adults with diabetes</th>
<th>2012–2030</th>
</tr>
</thead>
<tbody>
<tr>
<td>World</td>
<td>128,167.0</td>
<td>34.5</td>
<td>134,614.0</td>
<td>41.0</td>
<td>97,887.2</td>
<td>97,887.2</td>
<td></td>
</tr>
<tr>
<td>1 China</td>
<td>32,959.8</td>
<td>35.7</td>
<td>36,554</td>
<td>50.0</td>
<td>31,738.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 India</td>
<td>14,741.5</td>
<td>23.4</td>
<td>15,430</td>
<td>28.2</td>
<td>13,793.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 USA</td>
<td>10,069.6</td>
<td>41.8</td>
<td>10,206</td>
<td>52.0</td>
<td>5331.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Brazil</td>
<td>5142.9</td>
<td>38.5</td>
<td>4810</td>
<td>48.7</td>
<td>4398.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Mexico</td>
<td>2775.3</td>
<td>26.2</td>
<td>3290</td>
<td>35.2</td>
<td>3007.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 Bangladesh</td>
<td>940.2</td>
<td>17.0</td>
<td>888</td>
<td>21.9</td>
<td>2750.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 Indonesia</td>
<td>2022.2</td>
<td>26.8</td>
<td>2231</td>
<td>36.6</td>
<td>2299.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 Egypt</td>
<td>2104.8</td>
<td>27.9</td>
<td>2182</td>
<td>32.3</td>
<td>1896.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 Russia</td>
<td>5812.1</td>
<td>45.8</td>
<td>5096</td>
<td>53.9</td>
<td>1787.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 Japan</td>
<td>4350.4</td>
<td>61.2</td>
<td>4451</td>
<td>60.2</td>
<td>1756.9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Florez 2013 [1].

Aging of the overall population is acknowledged as one driver of the diabetes epidemic [9] but other risk factors such as obesity, reduced physical activity, low education, smoking, dyslipidemia, and hypertension have been reported across different populations [8,10–13]. Similarly the prevalence of prediabetes (impaired glucose regulation) increased with age, reaching up to 25% (men) and 30% (women) above 70 years of age [4].

Diabetes in older adults is associated with higher mortality, reduced functional status, increased risk of institutionalization, as well as an enormous human and economic burden [14]. There is also substantial risk for both macrovascular and microvascular complications associated with diabetes in the elderly. Major lower extremity amputations [15], myocardial infarction, visual impairment, and end-stage renal disease have higher rates in the elderly, particularly in those aged 75 years and older, than in any other age group. Although death rates related to hyperglycemic crisis have declined markedly in the past two decades, they are still higher in older adults with diabetes. In contrast there is a growing concern, particularly in those aged 75 years and older, with increasing rates of emergency department visits related to hypoglycemia [16].

Older adults with diabetes may either have incident disease or long-standing diabetes with onset in middle age or earlier. Demographic and clinical characteristics of these two groups differ in a number of ways, adding to the complexity of making generalized treatment recommendations for older patients with diabetes.

Pathophysiology of diabetes in the elderly

Many of the differences between older patients compared to younger patients with diabetes stem from the pathophysiology of aging combined with alteration in glucose metabolism including insulin secretion, insulin resistance, and hepatic glucose metabolism [17]. In addition, conditions associated with aging such as coexisting illnesses, increased adiposity, decreased physical activity, sarcopenia, and effect of medications used to treat comorbidities further impact abnormal glucose metabolism in the elderly [18].

Metabolic alterations

Abnormal insulin secretion and resistance to insulin action are primary pathophysiologic factors in development of diabetes in older adults. The abnormality in insulin secretion is related to alterations in insulin release, loss of pulsatility, loss of first-phase insulin secretion, and decreased response to incretin hormones with aging [19]. Abnormal insulin sensitivity is commonly seen with aging and is exacerbated by confounding factors such as higher body fat content and loss of fat-free mass [20]. Recent studies have shown that lean and obese older adults with diabetes have distinct metabolic profiles [21]. Compared to young adults, obese older people with diabetes have relatively normal insulin secretion, but marked resistance to insulin-mediated glucose disposal. On the other hand, lean older people with diabetes have profound impairment in glucose-induced insulin secretion, but minimal resistance to insulin action [21,22]. The older adult with diabetes also shows normal fasting hepatic glucose production compared to elevated hepatic glucose production in younger patients [22].

Environmental impact

Impact of diet and exercise is similar in older and younger populations. A highly processed, low-fiber, high-energy diet is linked to higher risk of diabetes at all ages. Aging causes structural and functional changes in skeletal muscles which forms the basis of changes in body composition and metabolic abnormalities such as insulin resistance [23]. Obesity, higher fat to muscle ratio, and lack of physical activity are common with aging and contribute to higher risk of diabetes.
Genetic predisposition
Genetic predisposition magnifies age-related changes in glucose metabolism although the specific genes responsible are not identified yet. A family history of diabetes and certain race and ethnic groups such as African American and Hispanic descent confer higher risk of developing diabetes with aging, supporting the hypothesis of genetic predisposition.

Other factors
A variety of aging-related changes have shown to add to the risk of abnormal glucose metabolism. Autoimmunity has shown to play a role in the impairment of glucose-induced insulin secretion in lean older people with diabetes [17]. Proinflammatory cytokines such as tumor necrosis factor alpha [24] and C-reactive protein [25] have been associated with higher risk of diabetes in elderly patients. On the other hand adiponectin was found to be associated with lower risk of diabetes in older men and women [26]. Age-associated reductions in mitochondrial oxidative and phosphorylation activity have also been shown to contribute to insulin resistance in the elderly [27]. Low levels of testosterone in men and higher levels of testosterone in women are associated with higher risk of insulin resistance and diabetes in the elderly [28].

Effect on clinical presentation
The unique pathophysiology of diabetes manifests as a unique clinical presentation of the disease in the older population. The renal threshold for glucose increases with age, leading to higher blood glucose level before glycosuria is detected [18]. As a result, urine dipsticks are less likely to detect abnormality and polyuria may not be a prominent symptom. Furthermore, polyuria may be absent because of impaired thirst mechanisms in older adults which also lead to dehydration. Thus, the classic symptoms of hyperglycemia—polyuria and polydipsia—are often not present. As a result, clinical presentation of diabetes is altered in older adults and may manifest as infections or failure to thrive leading to serious complications such as nonketotic hyperosmolar coma [29].

Prevention of type 2 diabetes
Numerous clinical trials have shown that in high-risk subjects (particularly those with impaired glucose tolerance or IGT), the incidence of T2DM can be prevented or delayed by lifestyle interventions or by various classes of medications which have been reviewed elsewhere [33]. The Da Qing Study (China) randomly allocated 33 clinics (557 persons with IGT) to 1 of 4 study conditions: control, diet, exercise, or diet plus exercise [34]. Compared with the control group, the incidence of diabetes was reduced in the three intervention groups by 31%, 46%, and 42%, respectively and with a modest weight loss in study participants. The Finnish Diabetes Prevention Study evaluated 522 obese persons with IGT randomly allocated on an individual basis to a control group or a lifestyle intervention group that emphasized physical activity, weight loss, limited total dietary intake and intake of saturated fat, and increased intake of dietary fiber [35]. During the trial, the incidence of diabetes was reduced by 58% in the lifestyle group compared with the control group. The U.S. Diabetes Prevention Program is the largest trial of primary prevention of diabetes to date and was conducted at 27 clinical centers with 3234 overweight and obese participants with IGT randomly allocated to 1 of 3 study conditions: control, use of metformin, or intensive lifestyle intervention [36]. The goal of lifestyle intervention was to achieve and maintain 7% or greater weight loss through a low-calorie, low-fat diet and 150 or more minutes of moderate physical activity weekly. Nearly 20% of study participants were age 60 years and older at enrollment. Overall, in 3 years the incidence of diabetes was reduced by 31% in the metformin group and 58% in the lifestyle group. Older adults had a greater benefit (71% reduction) with lifestyle intervention but did not appear to benefit from metformin [36,37]. Additional benefits of the lifestyle intervention that might impact older adults have been reported, including reduction in urinary incontinence [38], reduction in cardiovascular risk factors [33], and improvement in several quality-of-life domains [39]. Although these results suggest that diabetes prevention through lifestyle intervention may be pursued in relatively healthy older adults, the DPP did not enroll a significant number of study participants over the age of 70 years or those with functional or cognitive impairments. Preventive strategies that can be efficiently implemented in clinical settings and in the community have been developed and evaluated [40], but as yet there has been little focus on older adults in these translational studies.

Screening
Older adults are at high risk for both diabetes and prediabetes. Surveillance data in the US suggest that nearly half of older adults may have prediabetes [3] while data in Europe suggest that up to one third may be affected [4]. Screening for diabetes and prediabetes in older subjects should be in accordance with accepted guidelines but no age-modified criteria are currently recognized [30]. The American Diabetes Association recommends using either fasting plasma glucose, hemoglobin A1c, or oral glucose tolerance test (OGTT) [31]. In high-risk older subjects with a normal fasting glucose, and where an OGTT is not feasible, determination of HbA1c may be helpful [30]. The benefits of identification of prediabetes and asymptomatic diabetes in older adults depend on whether primary or secondary preventive interventions would likely be effective, and on the anticipated time frame of benefit of interventions weighted against the patient’s life expectancy [32].

Prevention and management of diabetes and its comorbidities in older adults

Screening
Older adults are at high risk for both diabetes and prediabetes. Surveillance data in the US suggest that nearly half of older adults may have prediabetes [3] while data in Europe suggest that up to one third may be affected [4]. Screening for diabetes and prediabetes in older subjects should be in accordance with accepted guidelines but no age-modified criteria are currently recognized [30]. The American Diabetes Association recommends using either fasting plasma glucose, hemoglobin A1c, or oral glucose tolerance test (OGTT) [31]. In high-risk older subjects with a normal fasting glucose, and where an OGTT is not feasible, determination of HbA1c may be helpful [30]. The benefits of identification of prediabetes and asymptomatic diabetes in older adults depend on whether primary or secondary preventive interventions would likely be effective, and on the anticipated time frame of benefit of interventions weighted against the patient’s life expectancy [32].
Challenges and management principles
Care for elderly patients with diabetes poses a unique clinical challenge [41]. The management of older patients with diabetes is complicated by the medical and functional heterogeneity of this group [17,42]. The heterogeneity of this population is a key consideration for clinicians developing intervention strategies and establishing targets for elderly patients with diabetes [30,32,43]. Clinical and functional heterogeneity include:

- Older adults who have just converted from prediabetes to diabetes, have a few comorbidities but no chronic complications and remain active with excellent functional status;
- Patients whose diabetes was recently diagnosed but have had significant hyperglycemia for many years and already have complications successfully compensated to preserve functionality;
- Elderly patients who have developed diabetes in middle age and developed over time multiple related comorbidities and complications, which have been partially compensated;
- Patients who are frail and even disable, with advanced cognitive impairment, multiple comorbidities and complications.

Since many older adults with diabetes are in-between the afore-mentioned stages, with mild or early functional limitations and multiple risks for worsening morbidity, they require an individualized plan based on their functional status and life expectancy [30,43,44]. Specific data on the treatment options for glycemic, lipid, and blood pressure control have been recently reviewed [32] and a framework for considering treatment goals in older adults was proposed. This framework is an attempt to balance the expected time frame of benefit of interventions with anticipated life expectancy [45,46]. Three major classes of older patients with diabetes and increasing levels of mortality risk were identified: (1) those who are relatively healthy; (2) those with complex medical histories where self-care may be difficult, and (3) those with a very significant comorbid illness and functional impairment [32].

The goal of physicians and other providers caring for elderly patients with diabetes should be to optimize glycemic control and reduce associated cardiovascular risk factors in an effort to maximize long-term quality of life [32,43,47]. On the other hand, for frail older patients, particularly those with severe comorbidities and disabilities, aggressive management is not likely to provide benefit and may even result in harm, as a consequence of frequent hypoglycemia associated with aggressive glycemic control [48,49]. Several drug therapy options alone and in combination are available for treatment of diabetes in the elderly [50]. At first glance, the therapeutic principles for glycemic control in elderly patients with diabetes are not very different from those in younger adults. However, the elderly with diabetes require additional areas of consideration in goal setting and use of treatment modalities. Treatment goals must be individualized, and at times the enthusiasm for tight glucose control may need to be weighed against safety concerns. In deciding the best strategy for the intervention, it is important to adjust the regimen and goals periodically as the diabetes progresses, complications develop and/or functionality may decline. In these patients, the regimens should be simplified with less frequent dosing, avoiding interaction with other drugs that may affect the treatment effectiveness. The presence of renal, liver and/or cardiovascular comorbidities may create contraindications and/or increase the risk of hypoglycemia. Other important considerations are cost and the effect of these medications on weight and/or lipids. In general, oral agents should be started at the lowest possible effective dose to minimize the risk of adverse events.

Geriatric syndromes in older adults with diabetes
As individuals age, they have a higher likelihood of developing multiple medical conditions, and with each of these conditions comes the probability that they will need to take multiple daily medications. In the past, the significant health burden in older people with diabetes was attributed primarily to higher risk of macrovascular and microvascular complications. However, many coexisting medical conditions (such as cognitive dysfunction, depression, functional disability, falls, polypharmacy, urinary incontinence, and chronic pain) not typically associated with diabetes occur at higher frequency in older adults with diabetes [32]. These conditions, termed geriatric syndromes, may interfere with a patient’s ability to perform self-care tasks, including glucose monitoring, understanding role of diet and exercise in glucose excursions, and following a complex insulin regimen. Difficulty with self-management may lead to increased risk of nonadherence or treatment errors, contributing to an increased risk of hypoglycemia and poor glycemic control, which in turn leads to further difficulty with self-management and ultimately increased risk of morbidity and mortality. Figure 57.1 shows the complex interaction seen in older patient with diabetes. Thus, before developing treatment and management plans for older adults, it is of key importance to conduct a careful and comprehensive assessment for all potential comorbidities, which will very possibly impact how you plan your therapeutic approach.

Cognitive impairment
Diabetes is associated with 1.5-fold greater risk of cognitive decline and 1.6-fold increased risk of future dementia [51]. Cognitive dysfunction in older adults with diabetes may manifest as deficits in psychomotor efficiency, global cognition, episodic memory, semantic memory, and working memory [52]. In particular, executive functioning which is mediated by the frontal lobe, affects behaviors such as problem solving, planning, organization, insight, reasoning, and attention. Many elderly patients with cognitive dysfunction remain undiagnosed; particularly in the early stages as subtle declines in cognitive function remain unrecognized by family as well
as healthcare providers. Cognitive dysfunction can interfere with self-care abilities such as following medical, nutritional, and exercise regimens, leading to increased risk of treatment complications. In a small study examining patients with diabetes over the age of 70 years, executive dysfunction was identified in more than one third of patients and was associated with poor glycemic control [53]. In clinical practice, short screening tools such as a clock drawing test are useful to assess executive function. An important aspect of recognizing cognitive dysfunction is to realize that these people may not recognize hypoglycemic episodes, may not be able to treat them appropriately, and may forget to report them to providers. Thus they are at high risk of frequent unrecognized hypoglycemic episodes. It is important to provide diabetes education to family members or other caregivers to ensure the safety of the patient.

**Depression**
Depression is approximately twice as common in people with diabetes compared to the general population [54]. Diabetes and depression have a bidirectional relationship of increased risk [55]. Meta-analyses have demonstrated that diabetes doubles the risk of depression and that depression is associated with poor glycemic control [56]. Presence of depression is associated with poor adherence to medication and diet regimens, a reduction in quality of life, and an increase in healthcare expenditures [54,57]. A combination of diabetes and major depression leads to higher odds of functional disability compared to either diabetes or depression alone [58]. Depression is frequently undiagnosed and remains untreated in this high-risk population. Timely identification and treatment of depression may increase adherence and improve glycemic control along with overall quality of life for older adults.

**Polypharmacy**
This is a complex, challenging aspect of caring for elderly adults who often require multiple medications to optimally manage their diabetes and associated conditions [59]. Although, multiple medications are unavoidable in many, careful attention to appropriate dosing, avoidance of drug–drug interactions, interventions to increase regimen adherence, and assessment of financial feasibility are important. In order to treat more
effectively, it is important to keep the medication list current and review it at each medical visit.

**Physical disability and falls**

Older adults with diabetes have a reduction in physical function and overall health status compared to age-matched controls, and are more likely to use a mobility aid such as a cane or a walker [60]. In a large epidemiologic study, the incidence of disability, defined as onset of inability to do one or more functional tasks (walking 0.25 mile, climbing 10 steps, performing household chores, shopping, and cooking meals) was about twice as high for women with diabetes as those without [61]. Individuals with these disabilities may require assistance from other caregivers or family members to implement tasks such as self-monitoring of blood glucose (SMBG), giving insulin injections, meal preparation, and following a physical activity regimen. Falls, especially the ones resulting in injury, are also more common in older adults with diabetes [62,63]. The increased risk of falls in this population is multifactorial. Presence of peripheral and/or autonomic neuropathy, cardiovascular disease, reduced renal function, muscle weakness, lower limb dysfunction, functional disability, loss of vision, polypharmacy, and comorbidities like osteoarthritis may contribute to, and worsen the impact of, falls in frail older adults [64]. While good glycemic control prevents progression of some diabetes complications and therefore may decrease the risk of falls, hypoglycemia that occurs as a result of tight glycemic control may increase the risk of falls, underscoring the need for a balanced approach in older adults [65]. People who develop a fear of falling are usually reluctant to follow physical activity recommendations. Referral to an exercise physiologist and a supervised program for physical activity may help such people reduce their fear of falling, and thus lead to increased physical training and lower risk of falls.

**Urinary incontinence**

Diabetes is associated with higher risk of developing urinary incontinence. This condition in older patients with diabetes, especially women, is also multifactorial in etiology. Risk factors include urinary tract infection, vaginal infection, and autonomic neuropathy (resulting in either neurogenic bladder or fecal impaction) [29]. These conditions can be exacerbated by uncontrolled diabetes causing polyuria. Urinary incontinence frequently remains undiagnosed because women do not volunteer the information. Although there is no direct evidence to suggest deleterious effect of incontinence on diabetes control, identification and treatment are recommended to improve quality of life in women.

**Chronic pain**

Pain is considered the 5th vital sign and should be evaluated at each visit. Many older adults are reluctant to complain about pain. However, pain can increase depression and impair ability to perform self-care and exercise. Pain management may improve overall quality of life in addition to diabetes care.

**Diabetes education, self-management, and decision support in the elderly**

**Diabetes self-management**

Self-management relates to tasks that an individual must undertake to live well with one or more chronic condition [66]. It is based on the theory that greater patient confidence in his/her capacity to make life-improving changes yields better clinical outcomes. The goal of self-management education is to increase self-efficacy and improve clinical outcomes by increasing patients’ confidence in medical management, coping skills, and skills for health and wellness [66]. The rationale for chronic disease self-management programs is to ultimately change a patient’s behavior by increasing the patient’s self-efficacy and condition-specific knowledge [67]. Studies targeting self-management in diabetes mellitus have found it to be effective. A study of a self-management program for Spanish-speaking diabetics demonstrated significant improvements over usual care in HbA1c, symptoms, health status, and self-efficacy at 6 months with improvements persisting at 18 months [68]. A similar community-based self-management program showed improvements in health behavior, self-efficacy, and health status among Hispanic older diabetics at 4 months and 1 year [69]. A community-based self-management program intervention failed to show any improvements in HbA1c or healthcare utilization in diabetics over usual care. There were, however, significant improvements in patient activation, self-efficacy, hypoglycemia, depression, communication with physicians, and adherence to healthy behaviors [70].

**Internet-based self-management interventions**

In recent years these interventions have become widely available. These programs offer patients the freedom to use them at their own pace, and in the comfort of their homes instead of at doctors’ offices or at other healthcare institutions. This may represent a significant advantage for older adults, including those who are unwilling or unable to go to healthcare facilities due to mobility impairment or problems with transportation. A few studies have explored the use of Internet-based learning for healthcare of older adults. An Internet-based intervention to improve self-management behaviors and self-efficacy in older adults with diabetes mellitus through synchronous communication (instant messaging and chat) and asynchronous communication (email and a bulletin board) was effective in improving HbA1c levels after 6 months [71]. A large controlled trial randomized 761 diabetic participants to an Internet-based self-management program, and the same program with email follow-up reinforcement and usual care. Individuals in the two treatment groups demonstrated significant improvements in HbA1c, patient activation and self-efficacy at 6 months. The addition of email
reinforcement did not show an effect [72]. Most recently, online personal health records have emerged as a type of Electronic Health Record (EHR) tool, intended to increase patients’ access to and ownership over their healthcare information. Patients may be able to use an online personal health record as a means of providing self-management training and education [73].

**Clinical decision support in elderly diabetes care**

These are rule-based systems that provide information about drug interactions, drug-allergy problems, treatment algorithms, information about alternative medication regimens, and computer-based clinical pathways, or may have more advanced scoring and expert systems that assist clinicians by providing reliable and objective estimation of disease prognoses, probability of adverse events and outcomes [74]. Four features of clinical decision support systems are independent predictors of improved clinical practice: automatic provision of decision support as part of clinician workflow, provision of recommendations rather than just assessments, provision of decision support at the point of care, and computer-based decision support [75]. Evidence regarding the effects of clinical decision support comes from systematic reviews in varied clinical settings. A systematic review included seven randomized controlled trials comparing the effect of computerized clinical decision support versus usual care on diabetes processes of care and clinical outcomes. Five of the seven trials demonstrated significant improvements in clinicians’ performance on eye, foot, urine protein, blood pressure, and cholesterol examinations as a result of the intervention. Three out of the seven trials evaluating clinical outcomes failed to demonstrate the superiority of the intervention [74]. Most recently, a large controlled trial randomized patients to receiving either computerized clinical decision support or usual care. The intervention group physicians used the decision support system at outpatient visits by their patients with diabetes. As compared to the control group, the intervention group patients had significant improvements in HbA1c, systolic blood pressure control and maintenance of diastolic blood pressure control but no improvement in LDL–cholesterol [76]. A recent review of 33 studies showed the benefits of clinical decision support tools in improving evidence-based processes of diabetes care but a lack of effect on clinical outcomes [77].

**Shared decision-making aids**

These aids contribute to patient empowerment and better care with potential impact on healthcare outcomes. Despite their availability and demonstrated efficacy, their use is infrequent [78]. As medications are the most common treatment in diabetes mellitus, decisional aids may have a significant impact in the management of these patients. Paper-based decisional aids can enhance clinicians’ and patients’ discussions relating to diabetes medications during clinical encounters. Use of these decisional aids demonstrated acceptability, increased involvement in decision making and improved medication knowledge but did not affect adherence [79,80]. Decision aids can also be helpful in presenting risk information to diabetics. The format of presentation of risk information to diabetic patients represents a challenging task that could benefit from the use of shared decision-making aids [81,82]. Presenting cardiovascular risk in time frames shorter than 10 years improved the comprehension of cardiovascular risk and resulted in increased risk perceptions, emotional responses, and intent to change behavior in diabetics [82]. Patients with diabetes found natural frequency formats helpful for the presentation of risk information but most participants preferred graphical presentations in the form of bar charts aids [81]. More research is needed to ascertain the effectiveness, efficiency, and usability of decision aids for patients with diabetes.

Additional recommendations for the management of T2DM in older adults on a global scale were recently released by the IDF [83]. This guideline provides specific recommendations for education and diabetes self-management in older adults, including those who are functionally independent, those who are dependent (specifically addressing frailty and dementia), and end-of-life care.

**References**


CHAPTER 58

Psychological problems and management of patients with diabetes mellitus

Alan M. Delamater, Ashley Marchante and Jennifer Hernandez
Department of Pediatrics, University of Miami Miller School of Medicine, Miami, FL, USA

Key points

- Depression is common in patients with diabetes and has been shown to be associated with decreased self-care behaviors and poor glycemic control.
- Anxiety disorders are also common and have been shown to be associated with poor glycemic control.
- Eating disorders are common, especially in young women with diabetes, are persistent, and increase risk for metabolic control problems and microvascular complications.
- Neurocognitive problems can develop over time with history of poor glycemic control.
- Reliable and valid screening instruments with good clinical utility are available to identify patients with psychological problems and can be used in the clinical setting.
- There is some evidence to support the efficacy of psychological interventions incorporating cognitive behavior therapies to address psychological problems in patients with diabetes.
- Adoption of a patient-centered chronic care model of interdisciplinary care is a skillful approach to increase patient empowerment and improve clinical outcomes.

Introduction

This chapter considers psychological problems and their management in adult patients with diabetes. Both type 1 diabetes (T1DM) and type 2 diabetes (T2DM) are considered. The review is not exhaustive but rather focuses on methodologically sound research that examines psychological problems in patients with diabetes. We review selected studies illustrating the main psychological problems affecting adults with T1DM and T2DM, and also refer to recent reviews including meta-analyses. The main psychological issues include depression, anxiety, eating disorders, and compromised neurocognitive function. Within each, the prevalence, associated factors, and treatment-related issues are considered. The concluding sections discuss future research needs and recommendations for clinical practice.

Depression

Prevalence

Research has shown that adults with diabetes are much more likely to experience depression than adults without diabetes. The rates of depression in adults with both T1DM [1] and T2DM [2] are high in comparison to the general population. Evidence suggests a bidirectional relationship between depression and diabetes, such that the presence of diabetes increases the risk of developing depression and that the presence of depression increases the risk of developing diabetes [3]. Although the literature is unclear about the bidirectional relationship in T1DM, longitudinal research has found that individuals who experience high levels of depressive symptoms are at increased risk for developing T2DM [4]. Researchers suggest two hypotheses for this trend: behaviors associated with depression, such as poor self-care, may lead to obesity and insulin resistance, and biochemical changes in depression or treatment of depression could result in diabetes [3].

Associated factors

Depression in adults is associated with both medical and psychosocial risk factors, including poorer glycemic control [5], medical complications [1,6], lower health-related quality of life (HRQOL) [7], and greater nonadherence to medication [8]. Although research has found an association between diabetes-related depression and glycemic control, the nature of the relationship is still unclear. Some have found that diabetes-related depression is associated with higher glycosylated hemoglobin A1c (A1c) levels over time [5], but this may...
only be true for insulin users [6,9]. Depression is also associated with diabetes-related microvascular and macrovascular complications. In a study conducted over a 5-year period with patients who presented with major depression and diabetes, those with major depression had an increased risk of developing advanced microvascular complications, including end-stage renal disease and blindness, and macrovascular complications, such as heart attack or stroke [6]. The results of a recent meta-analysis indicated that depressed patients are 1.76 times more likely to be nonadherent than nondepressed patients with diabetes [8]. Thus, it is especially crucial to address regimen adherence in patients with diabetes who are depressed.

It is also important to consider psychosocial factors that accompany diabetes-related depression, as they may impact health outcomes as well. Researchers have found a strong association between HRQOL and depression in people with diabetes [7]. However, depressive symptoms have only been consistently associated with the mental components of HRQOL domains and less with the physical functioning domains [7]. Therefore, it is important to consider how specific domains of HRQOL vary across different groups of people who show symptoms of depression.

Treatments for depression
Due to the increased risk for poor health outcomes in individuals with comorbid depression and diabetes, it is important to consider diabetes-specific treatments. Evidence-based treatments for depression have been successful within the general population, including psychological interventions, antidepressants, and collaborative care. The results of a recent meta-analysis demonstrated that cognitive-behavioral therapy (CBT) was the most effective psychosocial intervention to decrease depression in patients with T1DM and T2DM; however, these interventions did not reliably result in significant improvements in glycemic control [10]. Future psychosocial interventions should include a self-care component to address glycemic control as part of CBT treatment for individuals with diabetes.

Antidepressants such as SSRIs have been effective in reducing depression in patients with diabetes, with trends towards improved glycemic control [11]. In a recent randomized, placebo-control trial, pharmacologic treatment with sertraline improved both depression and A1c in low-income minority patients with diabetes [12]. Collaborative care, which includes CBT and antidepressants, has been effective in reducing levels of depression. The results of a recent randomized control trial found that, in a sample of patients with depression and T2DM, an integrative care intervention was effective in reducing both A1c levels and depression [13]. This integrative care intervention included individualized programs to improve adherence to antidepressants and treatment adherence. While collaborative care treatment programs show promise for improving depression and glycemic control in adults with diabetes, future research should consider the long-term effectiveness of these interventions.

Anxiety
Prevalence
There is a substantial literature documenting the prevalence and course of both affective and anxiety disorders in adults with diabetes. For example, research findings have demonstrated that generalized anxiety disorder (GAD) is more common in patients with diabetes than in the general population [14]. The results of a meta-analysis indicated that 13.5% of patients with diabetes met criteria for GAD, 40% had elevated symptoms of anxiety, and 26.5% met criteria for having anxiety disorder not otherwise specified [14]. These researchers suggested that the higher rates of GAD may be due to the psychological stress of disease self-management or fears of hypoglycemia and medical complications. An early report indicated rates of disturbance due to depression and anxiety in a clinical sample of adults with diabetes as 41% and 49%, respectively, significantly higher than the 10–20% rates to be expected in the general population [15].

Associated factors
Gender is a significant correlate of both anxiety and depression in patients with T1DM and T2DM, with women exhibiting higher rates than men [15–17], particularly women of lower socioeconomic status [15]. Research also indicates that adults with diabetes who screen positive for anxiety are more likely to have poor glycemic control [15,17–19]. Anxiety is also greater for those patients who have diabetes-related health complications [15]. The results of a study in Ireland of 1456 patients with diabetes revealed 22% screened positive for significant symptoms of anxiety and depression, and those who did were more likely to have diabetes-related health complications, uncertainties about glycemic control, to be female, and to be smokers or have a history of alcohol abuse [17].

Treatments for anxiety
CBT is among the best studied intervention for anxiety disorders in the general population, and is considered to be a well-established treatment for panic disorder, agoraphobia, and GAD. However, there have been relatively few studies addressing the treatment of anxiety disorders in patients with diabetes, although there are some reports of the effects of interventions to reduce stress and improve coping skills.

Due to the possible deleterious effects of stress and anxiety on glycemic control, researchers have studied the effects on glycemic control of stress management including relaxation training in individuals with T2DM, demonstrating significant improvements in A1c over 1 year [20]. Various types of relaxation interventions have also been effective in improving glycemic control of patients with T2DM. For example, a recent study showed improved insulin resistance after a traditional Chinese relaxation exercise (Qigong) [21]. Several studies showed moderate improvements in state and trait anxiety scores for patients trained in relaxation [21–23].
A recent randomized controlled trial examined the effects of a mindfulness-based group intervention program in outpatient with diabetes. Results showed significant reductions in stress, anxiety, and depression, as well as improved HRQOL in patients receiving the intervention; however, there was no impact on A1c [24]. Another randomized controlled study demonstrated that a group CBT intervention for adults resulted in reduced diabetes-related stress and improved coping and A1c, effects which were maintained at 6-month follow-up [25]. A recently reported study indicated that the positive effects of CBT for patients with anxiety and depression may be delivered over the Internet [26]. In this study, patients receiving the web-based intervention had significant reductions in diabetes-related distress and depression, but there were no effects on A1c [26].

Anxious individuals with diabetes may be susceptible to significant fears of hypoglycemia, injections, and/or diabetes-related complications. Research indicates that these fears are associated with poor adherence [27], highlighting the importance of interventions to reduce anxiety symptoms for these patients. There is evidence that for clinically anxious individuals with diabetes, CBT can be effective in reducing panic attacks, avoidance behaviors, and fear of hypoglycemia [28]. It is important to note that pharmacologic intervention including benzodiazepines and SSRIs can be effective in the treatment of anxiety disorders, but have not been well studied in adults with diabetes [29].

Eating disorders

Prevalence

Early studies indicated that eating disorders such as bulimia are common in adults with diabetes, especially among young women with T1DM [30]. Prevalence rates for anorexia, bulimia, binge-eating disorder, and eating disorders NOS have been shown to range from 5.4% to 7% in women with T1DM (lifetime 10.5% to 14.4%) and 6.5% to 9% in women with T2DM (lifetime 10% to 13.7%) [31]. In men and women, the prevalence of co-occurrence of binge-eating disorder and T2DM diabetes ranges from 2.5% to 26% [32,33]. The prevalence of subclinical levels of binge-eating patterns co-occurring with T2DM have been reported from 14% to 40% [34,35].

Rapid weight gain associated with the start of insulin treatment and the emphasis of dietary restrictions in diabetes management may help explain the increased prevalence of eating disorders in young women with T1DM. Moreover, insulin injection omission or under-dosing provides patients with T1DM a unique tool to inappropriately manage weight and therefore may partially explain the increased prevalence [36]. Rates of insulin omission among young adolescent girls with T1DM have been reported in the range of 11–14%, and up to 34% for older adolescents and young adult women [37].

Research indicates that there is a difference in the initiation of eating disorders between people with T1DM and T2DM: individuals with T1DM were more likely to develop an eating disorder after the onset of diabetes, while those with T2DM were more likely to have been diagnosed with binge-eating disorder (BED) above all other eating disorders, and more often, their eating disorder preceded their onset of diabetes [31].

Associated factors

Research has shown that patients with T1DM who have eating disorders and omit insulin are more likely to have poor glycemic control [37,38]. Prospective studies have shown that eating disorders in young women with T1DM are persistent and associated with an increased risk for microvascular complications including retinopathy, especially with continued insulin omission [39–41]. Another study indicated that insulin omission was associated with poor regimen adherence and higher rates of neuropathy than individuals who did not omit insulin [42]. Studies of young women who omit insulin indicate they have disturbed eating attitudes and negative attitudes about diabetes, report previous symptoms of anorexia and bulimia, and report that they are not truthful with their physicians about self-management [43].

Several studies have examined eating disorders in patients with T2DM. Results indicate that patients with binge-eating disorder report greater difficulties with hunger, greater impairments to quality of life, disinhibited eating (often at night-time), have higher rates of anxiety disorder, have higher rates of body image disturbance, are more likely to be overweight, and also report elevated symptoms of depression [33,44,45]. In addition, binge eating in women with T2DM has been associated with poorer psychological well-being, increased BMI, decreased self-efficacy for adherence to diet and physical activity recommendations, and poorer glycemic control [46].

Treatments for eating disorders

Treating eating disorders in patients with diabetes is obviously complicated by the role of diet in diabetes management. Treatment options can range from outpatient treatment, partial hospitalization, and inpatient hospitalization depending on the severity of the eating disorder. However, there is a limited evidence base for the treatment of eating disorders in patients with diabetes. In the general population, psychological treatment such as CBT has empirical support for the treatment of binge-eating disorder and bulimia nervosa, with some support for CBT and family therapy in the treatment of anorexia nervosa [47]. Psychopharmacologic interventions including SSRIs have also been useful in the treatment of eating disorders in the general population (mainly because of the association of eating disorders with depression and anxiety) [48].

For women with T1DM and eating disorders, a group administered CBT-based psychoeducational intervention was evaluated in a randomized controlled study, and was shown to reduce disordered eating but without change in glycemic
control or insulin misuse [49]. In a nonrandomized study, a small sample of women with T1DM and bulimia received an integrated multicomponent inpatient treatment (including CBT and family therapy) and were compared with a group of women receiving usual care: 3 years after the inpatient intervention, these patients had reduced symptoms of bulimia as well as improved psychological functioning and glycemic control [50]. Group therapy may be effective for treating eating disorders in patients with T2DM. One controlled study demonstrated reduced symptoms of binge eating at 3-month follow-up, and reductions in binge eating were associated with improved glycemic control [51].

**Neurocognitive problems**

**Prevalence**

A recent meta-analysis has shown that rates of cognitive impairment are higher among adults with diabetes than without diabetes [52]. In fact, research indicates that patients with T2DM have nearly double the risk of developing cognitive impairment and dementia [53]. Results of studies comparing cognitive dysfunction in T1DM versus T2DM patients have revealed differences in the kinds of cognitive impairments each group experiences: patients with T1DM tend to experience more changes in psychomotor speed and motor speed, while patients with T2DM may experience greater abnormalities in learning and memory [54].

**Associated factors**

There are several factors associated with cognitive impairment in patients with diabetes. Neurocognitive impairment is most often associated with chronic hyperglycemia [54,55], severe and frequent hypoglycemia [56], presence of microvascular complications including retinopathy and nephropathy [57], and comorbid health conditions such as hypertension and obesity, as well as longer duration of diabetes [54]. There is also evidence that cognitive impairment can make diabetes management especially challenging. In a study of older adults with T2DM, as cognitive impairment worsened, adherence to diabetes self-management tasks also declined, particularly for adherence to dietary and physical activity components of the regimen [58].

**Treatment**

Some evidence suggests that improving glycemic control may be helpful for cognitive impairments in older patients with diabetes [54,57,59]. Due to the association between poor glycemic control and cognitive decline, intensive insulin therapy to improve glycemic control should be used to improve cognitive functioning in patients with T1DM; in patients with T2DM, targeting risk factors such as hypertension and dyslipidemia may also be important to reduce cognitive dysfunction [54].

**Future research needs**

There is a large evidence base documenting the importance of psychological factors in diabetes management. However, more longitudinal studies with larger study samples should focus on the identification of specific psychosocial factors that influence regimen adherence and glycemic control over time in adults with T1DM and T2DM. For example, longitudinal studies can determine mechanisms to account for how psychosocial factors and health behaviors affect health over time. Results from such studies can inform the development of more effective interventions for specific patient populations. Sample sizes should be sufficient to detect effects among various patient subgroups with evidence specific psychological problems, including: ethnic minorities, low socioeconomic status, the elderly, those who live alone or without social and family support, those with diabetes-related health complications.

More research is needed to assess how to improve self-management skills, empowerment, coping skills and stress management to promote long-term health and quality of life for these special patient populations. Theory-based intervention studies are needed to determine how to individualize (i.e., tailor) psychosocial therapies to improve regimen adherence, glycemic control, and quality of life in the various patient populations.

Another research priority is to conduct multicenter clinical trials to document the efficacy of initial findings drawn from smaller, single-site intervention studies. Research should be conducted that evaluates the integration of psychosocial intervention into disease management program, for example by targeting high-risk patients. Cost-effectiveness trials in “real world” managed care settings are needed to assist in the translation of research findings to clinical practice. The effectiveness of behavioral and psychosocial interventions must be demonstrated in clinical settings and shown to be cost-effective in order for more widespread implementation of these therapies to occur.

**Recommendations for clinical practice**

This review has demonstrated that psychological problems in adults with diabetes are common in clinical practice, and that these problems are associated with decreased self-management [60]. Studies show that physicians are aware of these difficulties and how they may compromise optimal management of diabetes, yet they are often uncomfortable and frustrated by how best to deal with them [61]. There has been a paradigm shift in diabetes care in recent years, with the recognition that optimal diabetes management requires an appreciation of the context of healthcare, the interplay of multiple environmental, psychological, social, and medical factors, and the adoption of a patient-centered, chronic rather than acute model of healthcare delivery [62]. The importance of the relationship between...
physicians and other healthcare providers and their patients is critical [63], as is the provision of follow-up support over time [64].

The identification of psychological problems in clinical practice is a priority. Fortunately, there are a number of reliable and valid screening instruments that are brief, easy to score, and can be used in clinical practice. Depression in patients with diabetes is so common that has been referred to as “diapression” [65]. Clinicians can screen for depression using measures such as the Beck Depression Inventory-II [66], the seven-item Beck Depression Inventory-Primary Care scale [67], the Center for Epidemiological Studies-Depression scale [68], the Patient Health Questionnaire (PHQ-9) [69], or the Geriatric Depression Scale [70]. Many patients are also so stressed and anxious that their psychological distress impairs their ability to effectively manage their diabetes. The 20-item Problem Areas in Diabetes survey can be used to screen for diabetes-related psychosocial distress in patients with diabetes [71]. A more recent 17-item Diabetes Distress Scale can also be used to identify patients who feel overwhelmed with daily self-management [72]. Eating disorders in patients with diabetes—especially women—are prevalent, persistent, and lead to increased risks for health complications [73]. The Eating Disorder Examination-Questionnaire can be used to screen for the presence of eating disorders [74]. Other easy-to-use measures for screening are the Eating Attitudes Test [75] and the Eating Disorders Diagnostic Scale [76].

It is obviously critical to have skilled mental health professionals—knowledgeable about diabetes and its treatment—as part of the interdisciplinary healthcare team. In many clinical settings, however, there are financial barriers to having psychologists, social workers, and psychiatrists as part of the clinic team, so that making referrals is the most common clinical practice. Unfortunately, many patients do not follow-through with mental health referrals. Nevertheless, it is important for physicians to conduct screening for psychological problems using reliable and valid questionnaires such as those described earlier as a first step; patients who screen positive can then be counseled and referred directly to the mental health professional associated with the diabetes team. Given the complexity of assessment for neurocognitive difficulties, referral to a clinical psychologist or neuropsychologist must be made to adequately determine patients’ neurocognitive functioning, a more time-consuming and often costly evaluation.

Optimal diabetes self-management can be improved by addressing psychological problems in clinical practice, especially when clinicians work collaboratively with patients. Research has demonstrated that interventions to increase patient empowerment lead to improved patient outcomes in terms of psychological functioning, diabetes self-management behaviors, and glycemic control [77]. There is some evidence for the efficacy of cognitive-behavioral therapies in managing depression [13], anxiety [20,24,25], and eating disorders [49,51] in patients with diabetes. Interventions addressing intrinsic motivation, healthy coping, and problem-solving skills have also led to improved patient outcomes [78–80].

Conclusions

Psychological problems in patients with diabetes such as depression, anxiety, eating disorders, and neurocognitive difficulties, are common and associated with increased risk for poor patient outcomes including compromised quality of life, reduced self-management behaviors, poor glycemic control, and microvascular and macrovascular complications. It is critical to address these problems in clinical practice. Reliable and valid questionnaires with good clinical utility are available to screen for depression, anxiety and distress, and eating disorders. Psychological interventions utilizing cognitive-behavior therapy have improved patient outcomes. Inclusion of mental and behavioral health professionals as part of interdisciplinary treatment teams and adoption of the chronic care model of care recognizing patient autonomy is recommended in order to effectively manage psychological problems in clinical practice.

References


CHAPTER 59
Glycated hemoglobin, serum proteins, and other markers as tools for monitoring

Susan M. Gerber¹, Douglas F. Stickle², Intekhab Ahmed² and Serge A. Jabbour²
¹Cooper Medical School of Rowan University, Cooper University Health Care, Camden, NJ, USA
²Jefferson Medical College, Thomas Jefferson University, Philadelphia, PA, USA

Key points

• Technological advances have greatly improved our ability to monitor glycemic control in patients with diabetes. One of the most important clinical advances in modern diabetes care is the ability to estimate an integrated value of glucose excursions over several months, providing a useful window for glucose monitoring called the hemoglobin A1c (HbA1c) test.

• A physician using one of several available glycated hemoglobin assays for patient management needs to be familiar with what form of the modified protein has actually been measured. These considerations will affect the clinical implications of the test results.

• The Diabetes Control and Complications Trial (DCCT) demonstrated the relationship between elevated HbA1c and risk of development of diabetic complications, and provided the impetus for the standardization of HbA1c assays.

• Many studies, including the DCCT, support long-term control of diabetes with monitoring of glycemia using the HbA1c.

• Currently, sufficient standardization of the various HbA1c assays has been achieved such that the results can be applied to populations in a uniform and time-specific fashion. Thus HbA1c is recommended as a screening test for diabetes and for risk stratification by detecting asymptomatic persons with prediabetes and allowing for appropriate lifestyle interventions.

• HbA1c testing offers the advantages of a non-fasting test, increased stability prior to analysis, and fewer fluctuations from patient stressors or illness.

• The A1c-Derived Average Glucose Study demonstrated a linear relationship between HbA1c and estimated average glucose (eAG) eAG, showing that HbA1c is an accurate measurement of average blood glucose over time and supporting the interpretation of HbA1c in the more familiar units used for blood glucose—the same units that patients use for self-monitoring of their blood glucose (SMBG).

• Fructosamine assays may be used in clinical settings to detect short- to medium-term changes in mean blood glucose. Fructosamine assays also assist in certain clinical conditions where measurement of HbA1c is considered less reliable.

• 1,5-Anhydroglucitol testing is a marker of glycemia that responds more readily to glucose trends than HbA1c, and also offers the advantages of a simple blood test that is metabolically stable and shows small biological variability.

• Protein glycation products undergo rearrangements towards the development of advanced glycosylation end-products, which have been associated with chronic complications of diabetes, including tissue damage and cardiovascular disease.

• The continuous glucose monitoring (CGM) device has made possible the concept of the “artificial pancreas” when linked with insulin pump technology for continuous subcutaneous insulin infusion controlled by a computer-run insulin delivery algorithm. Use of this closed-loop system to eliminate SMBG, monitoring with surrogate markers such as HbA1c, and self-adjustment of insulin may enable diabetic patients to reach previously elusive near-normal HbA1c levels with fewer extremes of hypo- and hyperglycemia and diabetic complications.

Introduction

Technological advances have greatly improved our ability to monitor glycemic control in patients with diabetes. The wide availability of relatively inexpensive handheld meters, for example, has facilitated the monitoring of spontaneous blood glucose levels at various times during the day. However, the notorious fluctuation of postprandial glucose values is a common source of frustration in patients with diabetes, as well as the physicians and nurses trying to determine the level of their glycemic control over a longer time frame. To this end, one of the most important clinical advances that has improved modern diabetes care is the recognition that glucose undergoes nonenzymatic chemical reactions with various proteins in
the blood, including hemoglobin, in a manner that allows the estimation of an integrated value of the glucose excursions over a period of weeks to months, providing a useful window for glucose monitoring using the hemoglobin A1c (HbA1c) [1,2]. The early 1980s also saw the emergence of a second group of tests, collectively called “fructosamine,” that were based on a determination of glycated serum proteins, especially albumin, and found to be useful for similar clinical purposes over a shorter time period than that for the HbA1c [3,4]. There has also been growing interest in utilizing measurement of advanced glycosylation end-products (AGEs) in serum and in tissues as laboratory correlates of some of the chronic tissue damage observed in patients with diabetes [5]. In the present chapter, starting from a chemical description of the glycation reactions, we will present the pathophysiologic background necessary for an understanding of the variety of tests now used for glycated hemoglobin (GHb) and albumin determination and their interpretation in different clinical encounters. We will discuss the use of anhydroglucitol as a monitor of glucose. We will provide an overview of the standardization of HbA1c measurement and reporting, as well as the use of HbA1c for diabetes screening and risk category stratification. We will briefly describe AGEs and their potential relationship to the chronic complications of diabetes. Finally, we will examine future directions of diabetes care.

**Glycation reactions**

During the early part of the twentieth century, Amadori, Maillard, and others characterized the fundamental chemical reactions that underlie the formation of HbA1c and more advanced protein products associated with aging and tissue damage [6,7]. Chemically, the basic reaction involves covalent attachment of the aldehyde group of an acyclic glucose molecule to protein amino groups (Figure 59.1). The first step of this reaction, resulting in the formation of a Schiff base, can be transient and reversible, while the second step yields a fixed ketoamine product. The potential clinical utility of these observations stems from the appreciation that these slow, nonenzymatic processes occur in vivo, and that the accumulation of protein glycation products is significantly increased in patients with hyperglycemia [8,9]. The best known of these products is GHb and its related derivatives, including the HbA1c subfraction, which has evolved over the past several decades from an intriguing finding in diabetic patients [10] to become a critically important clinical variable to evaluate the level of glycemic control in patients with diabetes over time [11–13]. In addition, protein glycation products undergo further rearrangements towards the development of AGEs, which have been associated with chronic complications of diabetes, including tissue damage and cardiovascular disease, as discussed later [4,5].

In solution, the native glucose molecule exists primarily as an inert closed-ring structure. The open form, constituting 0.001% of the dissolved glucose, however, has a reactive aldehyde group with a unique capacity to bind covalently with proteins, generally to the free ε-amino group of lysine side chains [14]. In hemoglobin A (HbA), the amino group of the β-chain N-terminal valine is also highly reactive with glucose [15,16]. In both instances, the initial product is the unstable Schiff base or aldime linkage, which may dissociate or spontaneously progress to a ketoamine or Amadori product, which is only partially reversible [17] (Figure 59.1). The rate of formation of the Schiff base depends predominantly on the ambient glucose concentration, while the formation of ketoamine is influenced by both the glucose concentration and the time of exposure of the protein to glucose [18–20].

**Glycation of hemoglobin**

The glycation of HbA, leading to the formation of HbA1c and other derivatives, is the best characterized of the clinically important protein glycation reactions, and is unique in several respects. One is the novel formation of a linkage between glucose and the amino group of the N-terminal valine of the hemoglobin β chain, which serves as a unique recognition target for clinical biochemical assay methods, as described later [21]. Secondly, in contrast to the glycation of other serum proteins, it is crucial to recognize that the glycation of HbA is an intracellular process. Thus, determinants of the concentration of glucose within the red cells as well as factors that affect the turnover and lifespan of the erythrocytes will strongly influence the level of GHb.

Glucose first enters red blood cells by facilitated diffusion through membrane transport proteins with a Michaelis constant of about 0.5–2.0 mmol L⁻¹ [22]. The intracellular glucose is phosphorylated by hexokinase and enters the three major pathways of erythrocyte glucose metabolism, including glycolysis to pyruvate/lactate, formation of 2,3-bisphosphoglycerate, and the pentose phosphate pathway. The activity of hexokinase and all three major metabolic pathways declines exponentially during the life of the red cell, with a major loss of enzyme activities during the transition from reticulocyte to mature cell [23]. The loss of intracellular capacity to metabolize glucose causes a more rapid equilibration of extracellular glucose within the intracellular compartment and can lead to a subsequent rise in the glycation of hemoglobin in aging red cells. Further support for this hypothesis comes from the finding that HbA1c concentrations are increased in nondiabetic patients with altered erythrocyte glucose metabolism such as those with glucose-6-phosphate dehydrogenase deficiency [24]. Interestingly, GHb is virtually undetectable in porcine red cells, since glucose transport across the pig erythrocyte membrane is very limited, and this minimizes the intracellular content of free glucose [25].

As pointed out earlier, glycation may also occur at other sites on hemoglobin, which may be detected or missed by some
of the routinely used methods for measuring GHB [14,15]. Thus, ion-exchange chromatography, operating on the basis of a charge difference, separates hemoglobin modified at the N-terminal valine, while weak acid hydrolysis detects all the glycation sites of HbA that occur with varying degrees of reactivity (see later). Therefore, a physician using one of several available GHB assays for clinical patient management needs to be familiar with what form of the modified protein has actually been measured. These considerations will affect the clinical implications of the test results [2,26].

Nomenclature

Glycosylation should be reserved to describe the enzymatic transfer of carbohydrates to proteins, such as during protein biosynthesis and posttranslational processing. In contrast, glycation is the term that best describes the process of nonenzymatic bonding of glucose to proteins, as in the formation of HbA1c. The most commonly employed tests for GHB in current clinical practice involve the following glycated protein derivatives:

1 Hemoglobin A (HbA) is “native” hemoglobin consisting of the unmodified globin tetramer formed from two α- and two β-chains.

2 Total glycated hemoglobin (GHB) includes glycation at the N-terminal valine of β-globin in HbA (the actual HbA1c component) as well as additional derivatives including glycation at other valines and lysines in the hemoglobin molecule. GHB also encompasses glycation of other hemoglobin types (HbS, HbC, HbF, and so on).

3 Hemoglobin A1 (HbA1) refers to all the glycated species derived from adducts to the N-terminal valine in the β chain of HbA. It consists of several components classified according to the nature of the carbohydrate moiety. For example, fructose-1,6-diphosphate forms HbA1a1, glucose-6-phosphate forms HbA1a2, and so on.

4 Hemoglobin A1c (HbA1c) is a component of HbA1 formed by nonenzymatic addition of glucose to the N-terminal valine of the β-chain of HbA. As it is a subfraction of GHB and HbA1, HbA1c content is always a smaller percentage compared to the other two.

5 Hemoglobin A0 (HbA0) is a major component of HbA that is posttranslationally modified, including addition of sugars but without an N-terminal valine adduct to the β chain.

Glycation in individuals with normal glucose tolerance

Kinetics of glycation

Studies in normoglycemic individuals with radiolabeled iron documented that HbA1c increases linearly during the lifespan of the erythrocyte [18], with the lowest HbA1c content in the youngest population of erythrocytes and the highest content in the oldest cells [27]. In patients with diabetes, this relation is less clear, reflecting day-to-day changes in the serum glycemic profiles [27]. At any rate, there is no evidence that there is a certain period of the erythrocyte lifespan when glycation occurs selectively. From these considerations it also follows that conditions causing a reduced lifespan of the erythrocytes will cause a reduction in measured HbA1c concentration [21]. In fact, it has been proposed that measurement of HbA1c may serve as a measure of the erythrocyte lifespan in individuals with normal glucose tolerance and without a hemoglobinopathy [28].

Glycated hemoglobin in nondiabetic populations

The range of HbA1c values varies significantly in the normal population. Several groups of investigators have attempted to determine the factors associated with the distribution of HbA1c in nondiabetic individuals. From these studies it is clear that variation in the erythrocyte lifespan does not seem to be the sole cause of variation in normal HbA1c values [29,30]. In the large, cross-sectional “France Telecom” study, involving 3240 subjects without known diabetes, pregnancy, or hemoglobinopathies, HbA1c rose with deterioration in glucose tolerance, age, body mass index (BMI), family history of diabetes, as well as in women with a history of delivering large babies, and after menopause [29]. No difference in the HbA1c was noted between males and females, and in both sexes HbA1c was higher in smokers than in nonsmokers. Although some
seasonal variation was observed, it did not assume a consistent pattern over a 2-year period of observation. In a stepwise multiple regression analysis, the most significant factors associated with HbA1c level were age followed by fasting plasma glucose (FPG) [29]. Even so, the strongest correlation (i.e., between HbA1c and FPG) was strikingly low ($r = 0.20$). Yudkin et al. [30] identified among nondiabetic individuals those who had relatively elevated or depressed GHb values, as confirmed by four different methods. The low and high glycators represented approximately 20% of the study population. The observed variation in GHb could not be explained on the basis of glucose level, age, gender, BMI, hemoglobin levels, smoking, dietary habits, or vitamin intake.

**Standardization of the HbA1c assay**

Because some of the variation in HbA1c levels in normal populations may be poorly explained by glyceremia, and because HbA1c testing was previously poorly standardized over the years, it had been difficult to recommend the use of these tests as a method of screening for diabetes mellitus or for monitoring treatment of diabetes in practice [31,32]. However, the Diabetes Control and Complications Trial (DCCT) published in 1993 demonstrated the relationship between elevated HbA1c and risk of development of diabetic complications, and provided the impetus for the standardization of HbA1c assays [33]. In the 1990s, the United States (US), Japan, and Sweden were utilizing a variety of HbA1c analyses with varying results from different laboratories [34]. The American Association for Clinical Chemistry formed a workgroup in 1993 to help standardize the HbA1c assays to make the results directly traceable to the DCCT. This workgroup produced an HbA1c standardization protocol and subsequently dissolved. In its place, in 1996, the US formed the National Glycohemoglobin Standardization Program (NGSP). The NGSP reference method uses ion-exchange high-performance liquid chromatography (HPLC), and laboratories and manufacturers of testing equipment can be certified through the NGSP if their instruments are calibrated to match results from the NGSP method [35]. Sweden and Japan also began their own analogous HbA1c standardization programs in the 1990s using direct comparison methods [34]. The methods used in these countries were not traceable to the NGSP, creating a lack of international standardization. Further, the NGSP reference method also measured other constituents in addition to HbA1c, making its methodology not metrologically sound [35].

As a result, in 1995, the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) formed a Working Group on Standardization of HbA1c to address the differences in HbA1c results reported by these countries' standardization programs. The goal was to find a specific assay solely for HbA1c that would be accepted as an international standard with metrological traceability. The IFCC did successfully develop a specific HbA1c assay using mass spectrometry and capillary electrophoresis that has been approved by all its member societies and is utilized in a worldwide network of reference laboratories [35]. This has provided a global anchor for metrological traceability using the IFCC reference measurement system and has thus improved intra-assay and inter-assay coefficients of variation in HbA1c results [36]. However, because of the complicated process and expensive equipment needed, clinical laboratories can only afford to use the IFCC reference method to calibrate their instruments before processing routine samples using other HbA1c assays. Additionally, because the IFCC HbA1c assay is specific, the normal range for HbA1c using the IFCC reference method is approximately 2% less than previous normal ranges and must be expressed in mmol A1c/mol hemoglobin in order to be metrologically correct [35].

The IFCC recommended that a percentage not be used to convey HbA1c results as this was not metrologically sound. Given this recommendation and the potential for ensuing confusion, another workgroup was convened to give recommendations for HbA1c reporting. In 2007, the IFCC, the American Diabetes Association (ADA), the International Diabetes Federation (IDF), and the European Association for the Study of Diabetes (EASD) put forth a consensus statement recommending reporting of HbA1c in IFCC units, NGSP units, and adding a third requirement for reporting called estimated average glucose (eAG, either in mg dl$^{-1}$ or mmol L$^{-1}$) if the designated targets of the A1c-Derived Average Glucose (ADAG) Study were reached [37]. This study was designed to demonstrate a direct relationship between HbA1c and average blood glucose such that the HbA1c results could be reported in mg dl$^{-1}$ or mmol L$^{-1}$ in deference to the initial IFCC recommendation against reporting HbA1c as a percentage. Previous studies, such as the DCCT, contained limited amounts of blood glucose data performed over a short period of time in a small number of participants, and were not intended to show the relationship between HbA1c and average blood glucose [35].

**Estimated average glucose (eAG) reporting**

The ADAG trial results were published after the consensus statement was released. An international multicenter trial sponsored by the ADA, EASD, and IDF, the ADAG trial determined the mathematical relationship and reliability of the correlation between HbA1c and eAG. The study included type 1 and type 2 diabetics and people without diabetes. Participants used continuous glucose monitors as well as traditional self-monitored blood glucose (SMBG), generating approximately 2700 glucose measurements per participant. The ADAG trial demonstrated a linear relationship between HbA1c and eAG and generated the following equation—derived from population regression of mean blood glucose (MBG) on HbA1c— to describe this correlation: 

$$[28.7 \times \text{HbA1c (\%)}] - 46.7 = \text{eAG (mg dl}^{-1})$$

This trial showed that HbA1c is an accurate measurement of average blood glucose over time, and supported the interpretation of HbA1c in the more familiar units used for blood glucose—the same units that patients use for SMBG [38].
A second consensus meeting on standardization of HbA1c was convened at the IDF meeting in 2009, and that consensus statement, published in 2010, affirmed that the IFCC reference system is the only valid anchor for HbA1c standardization, and that HbA1c results should be dually reported in mmol/mol and derived NGSP units (%) using the IFCC-NGSP master equation \( \text{NGSP}(\%) = [0.09148 \times \text{IFCC}(\text{mmol mol}^{-1})] + 2.152 \). eAG was mentioned for possible use in discussion with patients but was not required for reporting [39]. The results of a third consensus meeting in December 2011 have not yet been published [40]. Concern about adopting IFCC units, especially with regard to physician and patient education, has made US agencies and laboratories hesitant to implement its use. For the immediate future, the US has no plans to require IFCC units for HbA1c reporting. IFCC unit adoption remains controversial and each country has individually decided whether to proceed with its usage [41].

**Mean blood glucose versus estimated average glucose**

Mean blood glucose (the average blood glucose expressed on a patient’s personal glucometer) may differ from eAG. One study found the eAG either over- or underestimated MBG by approximately 28 mg dL\(^{-1}\) in one third of patients, translating to an HbA1c difference of approximately 1% [42]. In patients with a higher HbA1c, the 95% confidence interval (CI) for the attendant eAG is wider, reflecting greater uncertainty. Also, because patients are more likely to check SMBG preprandially and at other times when blood glucose is often lower, the MBG is often lower than the eAG, which includes episodes of postprandial hyperglycemia. Patients should be made aware that these estimates are part of the range of the “true mean” of their blood glucose measurements [35].

**HbA1c as a screening test for diabetes**

Currently, sufficient standardization of the various HbA1c assays has been achieved such that the results can be applied to populations in a uniform and time-specific fashion. This finding led the International Expert Committee and the ADA in 2009 to recommend the use of HbA1c as a screening test for diabetes. The threshold HbA1c for diagnosis is 6.5% or higher, the cutpoint above which an increased prevalence of retinopathy is found. It is recommended that an HbA1c assay standardized or traceable to the DCCT reference assay and certified by the NGSP be used. Virtually all central laboratory methods in use in the US are certified to conform to the NGSP standard. It is not recommended to use point-of-care HbA1c assays for diagnostic purposes given their insufficient accuracy [43].

HbA1c testing offers the advantages of a non-fasting test, increased stability prior to analysis, and fewer fluctuations from patient stressors or illness. However, HbA1c testing is more expensive than fasting plasma glucose (FPG) and oral glucose tolerance testing (OGTT), and is not available everywhere in the world. Despite the lower sensitivity of HbA1c testing than FPG or OGTT, HbA1c testing may detect more cases of diabetes due to increased convenience. Additionally, there may be incomplete correlation between eAG and HbA1c in some individuals, and the possibility of increased glycation in some races/ethnic groups remains controversial. Finally, the epidemiologic studies on HbA1c from which recommendations of HbA1c cutpoints for diabetes diagnosis have been issued have all been performed in adults. Consequently, the optimal HbA1c cutpoints for diagnosis of T2DM in children remain uncertain [43].

In conditions that involve abnormal red blood cell turnover (such as some anemias, and recent transfusions or blood loss), HbA1c cannot be used for diabetes diagnosis. In the presence of hemoglobin variants, an HbA1c assay with which the abnormal hemoglobins do not interfere should be used for A1c analysis. Generally this means that A1c should be measured by a glycated hemoglobin assay which is in turn calibrated to A1c [43].

Screening for diabetes should begin no later than age 45 given that age is a major risk factor for the development of diabetes. Asymptomatic adults who are overweight and have at least one other risk factor such as hypertriglyceridemia, high-risk ethnicity (i.e., Pacific Islander) or race, sedentary lifestyle, hypertension, low HDL-cholesterol, PCOS, acanthosis nigricans, cardiovascular disease, previous impaired glucose tolerance (IGT) or impaired fasting glucose (IFG) or HbA1c ≥5.7%, history of delivering a baby heavier than 9 lbs, or history of gestational diabetes mellitus (GDM) should be screened for diabetes mellitus regardless of age. If the results are normal, testing should be repeated every 3 years. If testing is abnormal, screening should be repeated more frequently [43].

**HbA1c used for risk stratification**

The Expert Committee on Diagnosis and Classification of Diabetes Mellitus in 1993 and 2007 has recognized a category of increased risk for diabetes called prediabetes. Prediabetes is defined as an elevated blood glucose that—while abnormal—does not meet criteria for diabetes, but marks an individual who is at high risk for developing diabetes and cardiovascular disease in the future. The risk rises in a curvilinear fashion with increase in HbA1c. Prediabetes can be quantified via FPG or OGTT as well as HbA1c. In terms of HbA1c, prediabetes exists when HbA1c is between 5.7 and 6.4% [43]. A systematic review of 16 cohort studies involving approximately 44,000 persons over an average follow-up interval of 5.6 years showed that those with a HbA1c of 5.5–6.0% had a moderately increased relative risk of diabetes (9–25% 5-year incidence), and those with a HbA1c of 6.0–6.5% had a 5-year risk of developing diabetes between 25–50%, with a 20 times greater relative risk than persons with an HbA1c of 5.0% [44].

Diabetes is often diagnosed after complications appear. Appropriate screening at the above intervals can reduce potential glycemic burden and risk of attendant diabetes complications by detecting asymptomatic persons with prediabetes and allowing for appropriate lifestyle interventions.
HbA1c as a unique index of glycemic control in type 1 diabetes

In diabetic patients, the HbA1c can provide the physician with a valuable monitor of the adequacy of control of hyperglycemia. Interestingly, early studies showed that physician judgment about the quality of the antecedent glycemic control based on historical and laboratory data (glycosuria and fasting and random glucose tests) was frequently discordant with the actual data provided by measuring the GHb [45]. Knowing the HbA1c has gradually become a fundamental feature of modern diabetes care. As a time-averaged monitor of blood glucose, it prompts the physician to work with the patient to intensify management when necessary in order to achieve treatment goals [46]. Measurement of HbA1c is especially valuable in T1DM when day-by-day oscillations in glycemia tend to be highly variable, making an accurate assessment of the adequacy of diabetic treatment based on home blood glucose measurement very difficult. In this respect, if there is frequently a discrepancy between glucose values and the HbA1c, with few rare exceptions (discussed in more detail later), the latter is usually more correct [47].

The overall utility of the HbA1c in adjusting diabetes therapy is tempered by the fact that while it can help to identify patients with poorly controlled hyperglycemia, a high level by itself cannot guide the physician toward specific adjustments in the insulin regimen [48]. The HbA1c is thus an important ancillary test to be used in the ongoing assessment and treatment planning for the diabetic patient. Comparison of the HbA1c with the patient’s own capillary glucose testing enables the physician to alter therapy, especially the timing and amount of multiple insulin injections, to improve the overall diurnal glycemic control.

HbA1c testing for diabetes screening and monitoring in pregnancy

Previously, gestational diabetes mellitus was defined as any hyperglycemia or glucose intolerance diagnosed during pregnancy, whether or not the condition persisted postpartum or may have preceded the onset of pregnancy. Given the increased prevalence of T2DM in women of childbearing age, the current guidelines recommend screening those women with risk factors for T2DM at the first prenatal visit using standard diagnostic criteria. Women who meet diagnostic criteria at the first prenatal visit are diagnosed with overt diabetes, not gestational diabetes (GDM) [43].

Maternal hyperglycemia in early pregnancy can lead to an increased incidence of fetal malformations and wastage, and poor glycemic control during later pregnancy is associated with perinatal morbidity [49]. After the results of the international multicenter epidemiologic Hyperglycemia and Adverse Pregnancy Outcomes (HAPO) study [50] of ~25,000 pregnant women were published in 2008 demonstrating that risks of poor maternal, fetal and neonatal outcomes increased continuously with increasing maternal blood glucose at 24–28 weeks—even at glycemic ranges previously deemed normal during pregnancy—the diagnostic criteria for GDM were revised by the International Association of Diabetes and Pregnancy Study Groups (IADPSG), a consensus group convened in 2008–2009. All women who are not known to have overt diabetes are now screened between 24–28 weeks gestation for GDM using a 75-g OGTT. Additionally, new diagnostic criteria for the OGTT results were set based on odds ratios for adverse outcomes compared with the mean glucose levels of HAPO study participants [43]. The diagnosis of GDM is now made if any one of the following is found during OGTT: a FBG (after 8-hour fast) of ≥92 mg dL\(^{-1}\) (5.1 mmol L\(^{-1}\)), 1-hour BG of ≥180 mg dL\(^{-1}\) (10.0 mmol L\(^{-1}\)), and/or a 2-hour BG of ≥153 mg dL\(^{-1}\) (8.5 mmol L\(^{-1}\)). This is also a change in that only one abnormal value is now needed to make the diagnosis, which will increase prevalence of diagnoses of GDM. In 2011, the American College of Obstetrics and Gynecology recommended reverting to use of the previous World Health Organization (WHO) diagnostic criteria for GDM [43]. A systematic review of the WHO and the IADPSG criteria for GDM diagnosis published in 2012 showed that both criteria identified women at a small increased risk for adverse pregnancy outcomes, with associations of similar magnitude for each criteria, but with a high level of inconsistency for the newer IADPSG criteria [51]. Further studies are needed to better characterize the IADPSG criteria.

It is recommended that women with GDM undergo screening for underlying previously undiagnosed overt T2DM at 6–12 weeks postpartum, using nonpregnant OGTT diagnostic criteria. It is not recommended to use HbA1c for diagnosis of overt diabetes at that time because of the antecedent treatment for hyperglycemia [52]. Given their increased risk of development of overt diabetes [53], women with a history of GDM must be followed with periodic screening for prediabetes or diabetes [43]. There remains considerable confusion regarding the wide array of assays available to clinicians in the US, including testing conforming to both the ACOG and IADPSG diagnostic criteria. This reflects the ongoing debate between proponents of the HAPO study who reiterate the potential benefits of having a universal test to diagnose GDM, a common disease whose treatment offers great potential benefit to mothers and babies, and those who critique the HAPO study results as being merely observational and use of its diagnostic criteria as the cause of inappropriate over-diagnosis of GDM [54,55].

In pregnant patients with preexisting diabetes, the recognition that the risk of fetal malformations rises sharply with poor diabetic control prompted the use of HbA1c assays at the first prenatal visit to evaluate the patient’s glycemic control over the initial few weeks of gestation [56]. Conversely, the results of multiple nonrandomized trials showed significantly lower rates of congenital malformations in babies born to diabetic women who participated in preconception care programs with...
multidisciplinary support to train patients how to intensively manage their diabetes through diet, insulin use, and SMBG, with a goal of normalization of HbA1c prior to conception [57]. Ideally, the diabetic patient's blood glucose levels should be managed aggressively prior to conception to achieve a normal or near-normal HbA1c value and minimize the risk to the fetus early in gestation. However, almost two-thirds of pregnancies in diabetic women are not planned [43], and the early HbA1c value may be used in counseling the patient with regard to possible effects on the fetus [57,58]. Multiple medications used to treat diabetes and associated complications (such as statins) are contraindicated during pregnancy, and appropriate pregnancy prevention counseling must be provided to women of childbearing age prior to beginning these therapies. NPH, regular, lispro, aspart, and detemir insulins are pregnancy category B (no evidence of risk in humans). Glargine and glulisine insulins are category C, with insufficient data to establish safety in pregnancy. Metformin and acarbose are pregnancy category B, and all other oral antihyperglycemics are category C [43].

Frequent measurement of HbA1c is a useful adjunct to self-monitoring of blood glucose during a diabetic pregnancy. Since tight regulation of glycemic excursions is an important goal, HbA1c measured at 6-week intervals can help in the overall assessment of the mean blood glucose as estimated from fasting and postprandial capillary values [59]. As in nonpregnant patients with diabetes, the HbA1c value should corroborate the glucose testing results and lead to a reassessment of the diabetic regimen if poor control is suggested [60]. It is important to note here that the average blood glucose concentration is about 20% lower in normal pregnant women than in nonpregnant women, leading to a small reduction in HbA1c values; so the reference range for HbA1c during pregnancy should be 20% lower than in normal people [59,60].

**HbA1c and chronic complications of diabetes**

In long-lived proteins, once glycation has occurred, the protein modifications progress to more complex derivatives called AGEs. The involvement of AGEs in diabetic complications has been suggested by the observation that there is no significant correlation between the presence or severity of chronic complications of diabetes and the HbA1c level if the duration of observation is less than 2 years [61–64]. Conversely, observations exceeding the threshold of 2 years have revealed an association of elevated HbA1c with the development or progression of diabetic retinopathy, nephropathy, and neuropathy.

Early studies demonstrated that HbA1c was clearly predictive of the chronic complications of diabetes. In the seminal work by Klein and colleagues [65] on more than 1800 patients with diabetes, HbA1c at the initial examination was predictive of the 4-year incidence and progression of diabetic retinopathy. In the group with onset of diabetes at less than 30 years of age, who were all treated with insulin, those with HbA1c above the highest quintile compared with those below the lowest quintile had 1.9 times increased risk for developing any form of retinopathy; the relative risks for proliferative retinopathy and for the progression of retinopathy were 21.8 and 4.0, respectively. In older onset patients not treated with insulin, the same comparison yielded a relative risk for any form of retinopathy of 4.0 and a relative risk for progression of 6.2. Further early studies confirmed that over a period of 3–9 years, microalbuminuria was 3.6 times and retinopathy 2.5 times more prevalent in patients with mean HbA1c over 12.3% than in those with mean HbA1c less than 10.9% [66]. This group also showed that none of the patients developed retinopathy and microalbuminuria if their GHb was consistently below 9.0% (using a normal range of 6.3–8.2%).

The strongest data supporting that long-term control of diabetes with monitoring of glycemia using the HbA1c provides an index of the risk for development of microvascular diabetic complications comes from the DCCT [33]. This multicenter, 9-year study in a cohort of 1441 patients with T1DM compared conventional insulin therapy with diabetes management intensified by multiple daily insulin injections or use of an insulin infusion pump and rigorous attention to glycemic control. The intensive treatment resulted in significantly better glycemic control as estimated by HbA1c measurements, reduced from ~9% in the conventionally treated group to ~7% in the intensively treated group. In the primary prevention group (no signs of retinopathy at the start of the study) intensive insulin therapy reduced the onset of retinopathy by ~34%, and reduced microalbuminuria and clinical neuropathy by 35 and 60%, respectively. In addition, there appeared to be a continuous relationship between the HbA1c level and the risk of sustained progression of retinopathy in the entire study cohort. Consistent with these data is the notion that the higher incidence of chronic diabetic complications in the patients with elevated HbA1c was somehow causally related, perhaps through a similar mechanism involving protein glycation in target tissues.

A follow-up study to the DCCT, called the Epidemiology of Diabetes Intervention and Complications (EDIC), looked at the effects of intensive control on retinopathy, nephropathy, and neuropathy. It also examined incidence of cardiovascular (CV) events and the impact of intensive control on CV risk, and found that intensive blood glucose control reduced risk of any CV event by 42%, and reduced risk of nonfatal myocardial infarction, cerebrovascular accident, or death from CV cause by 57% [67].

In a large cohort of more than 5000 patients with T2DM, the United Kingdom Prospective Diabetes Study (UKPDS) similarly reported that intensive treatment of glycemia using lifestyle modification in addition to oral agents and insulin, resulting in a lowering of the group mean HbA1c from 7.9 to 7.1%, was associated with a marked reduction in microvascular complications of diabetes [68]. In the updated epidemiologic analysis of the UKPDS data, each 1% reduction in updated
mean HbA1c was associated with reductions in risk of 21% for any endpoint related to diabetes, 21% for deaths related to diabetes, 14% for myocardial infarction, and 37% for microvascular complications, with no threshold of risk observed for any endpoint [69]. Although not strictly proven in the intervention trial, these observational data suggest that in patients with T2DM, any reduction in HbA1c is associated with reduced risk of complications including CV events, even with values extending into the normal range (<6%) (Figure 59.2).

In the Norfolk cohort of the European Prospective Investigation of Cancer and Nutrition (EPIC-Norfolk), HbA1c was a predictor of death from cardiovascular and all causes in 4662 men aged 45–79 not known to be diabetic [70]. HbA1c was continuously related to subsequent all-cause, cardiovascular, and ischemic heart disease mortality through the whole population distribution, with lowest rates in those with HbA1c concentration below 5%. An increase of 1% in HbA1c was associated with a 28% increase in risk of death independent of age, blood pressure, serum cholesterol, BMI, and cigarette-smoking habit. These provocative data suggest that the level of HbA1c even within the normal range may be closely associated with excess mortality in a population of men not known to be diabetic.

The Action to Control Cardiovascular Risk in Diabetes (ACCORD) trial was a multicenter randomized trial comparing current diabetes practice guidelines with intensive blood glucose control, lipid reduction, and blood pressure control in 10,000 patients with established T2DM for an average of 10 years. All study participants also had high risk of future CV events due to either preexisting CV disease, evidence of subclinical CV disease, or two or more CV disease risk factors. All participants took part in the blood sugar control portion of the trial, and were additionally enrolled either in the blood pressure trial or the lipid trial. The results showed that decreasing blood pressure to below currently recommended levels did not decrease risk of fatal or nonfatal CV events. Also, treatment of study participants with statin plus fibrate did not decrease risk of CV events over statin alone. Goal HbA1c in the intensive blood glucose control group was <6.4%, achieved in half of the intensive control group. The National Heart, Lung, and Blood Institute discontinued the intensive blood glucose portion of the study in 2008 due to safety concerns, as there was a 22% increased rate of death in the intensive blood glucose control group, which was statistically significant. A trend was seen toward a lower rate of CV events, mostly nonfatal myocardial infarction, in the intensive BG control group. However, it was felt that the possible harm outweighed the potential benefit in this group. No specific cause for the higher death rate among the intensive blood sugar treatment group could be identified [71].

Assays of glycated hemoglobin

Different analytical procedures measure one or more of the glycated species of HbA denoted as HbA1, HbA1c, or GHb (Table 59.1). Generally, the currently available technology can be divided into two major groups. The first group consists of techniques where separation of glycated and nonglycated hemoglobin is based on the charge difference between HbA1 and HbA0, including ion-exchange chromatography, high performance liquid chromatography (HPLC), isoelectric focusing, and agar gel electrophoresis. The second group is based on the presence of glucose within the glycated species; thus, by definition, these techniques measure total GHb. A third group uses immunoassay techniques using new monoclonal

---

**Table 59.1 Assay methods for glycated hemoglobin and serum proteins**

- Method based on charge difference
  - Ion-exchange chromatography
  - High performance liquid chromatography (HPLC)
  - Isoelectric focusing
  - Agar gel electrophoresis
- Methods measuring total GHb
  - Weak-acid hydrolysis
  - Affinity chromatography
- Immunoassay-based methods for HbA1c
- Assays for glycated serum proteins
  - Fructosamine assay
  - Affinity chromatography
antibodies with idiotypes that specifically recognize the glucose adduct of the N-terminal valine of HbA in the HbA1c molecule.

As noted previously, the IFCC reference method has emerged, defining HbA1c as Hb that is irreversibly glycated at one or both N-terminal β-chain valines. To calibrate the reference method, after isolation from human blood, pure HbA1c and pure hemoglobin A0 are combined in predefined proportions from the certified primary reference material set that is used to calibrate the primary reference measurement system. Two reference methods measure the N-terminal β-chain peptides: in step one, hemoglobin is cleaved into peptides by the enzyme endoproteinase Glu-C; then in the second step, the β-chain glycated and nonglycated N-terminal peptides are separated and measured via HPLC. The reference method values are assigned to secondary reference materials (e.g. whole blood). On incubation with the enzyme endoproteinase Glu-C, the N-terminal hexapeptide of the β chain is cleaved off and glycated and nonglycated hexapeptides are separated and quantified by HPLC plus either electrospray ionization mass spectrometry or capillary electrophoresis with UV-detection (both give identical results). HbA1c is then measured as the ratio between the glycated and non-glycated N-terminal peptides, and is verified via calibration [34,39,72]. Studies comparing the results from the network of laboratories using the IFCC reference method show an ideal intra-laboratory coefficient of variation (CV) of 0.5–2% and inter-laboratory CV of 1.4–2.3% [34].

Methods based on charge difference
Ion-exchange chromatography and high performance liquid chromatography (HPLC)

Ion-exchange chromatography, the most commonly used technique, is essentially based on the method of Trivelli [73]. The original method was laborious and time-consuming, and required the use of large columns and volumes of elution buffers containing cyanides, as well as very stringent control of pH and temperature [74]. The newer assays utilize microm-columns and temperature control column jackets and offer the advantage of speed and relative ease of handling [75]. Kits have been commercially available for some time which contain prepackaged columns with prepared buffers, standards, and additives to eliminate Schiff bases (labile component). Commercial minicolumns are capable of measuring either HbA1 or HbA1c.

HPLC also uses a principle of charge separation of HbA1 or HbA1c from HbA0 but with much greater resolution and speed that lends the technique to automation. Some of the commercial HPLC systems have automatic samplers with a capacity to load up to 100 samples and an analysis time of only a few minutes per sample [76–78].

Isoelectric focusing and agar gel electrophoresis

Since the isoelectric difference between HbA1c and HbA0 is only 0.01–0.02 pH units, conventional electrophoresis is not capable of resolving HbA0 and HbA1c [79]. Isoelectric focusing using a thin-layer polyacrylamide gel containing ampholyte over a pH of 6–8 can resolve these species, but is prohibitively expensive for clinical use [80,81]. Of this group of techniques, electroendosmosis is more commonly used. This method employs agar gel plates with a citrate buffer; the gels are analyzed by scanning densitometry and are capable of differentiating HbA1 from HbA0 in the clinical laboratory setting [82,83].

Methods measuring total glycated hemoglobin
Weak-acid hydrolysis

This method relies on the ability of weak acids to release ketoamine from GHb and convert it to 5-hydroxymethyl furfural (5-HMF) [84]. The amount of 5-HMF can be determined colorimetrically after reaction with thiobarbituric acid. The initial methods had a number of disadvantages, including difficulty in achieving a complete hydrolysis, breakdown of 5-HMF, interference by glucose in the colorimetric assay, and other problems, but further technical modifications have been made to circumvent these shortcomings [85]. Although virtually every step in the method can be automated, the method is not widely used for clinical purposes.

Affinity chromatography

The principle of this method lies in the affinity of boronic acids for the cis-diol groups of glucose and on the reversibility of the reaction in aqueous solutions [86]. The commercially available systems based on this principle employ prepacked affinity columns that contain m-aminophenylboronic acid immobilized on either crosslinked beaded agarose [87,88] or Sepharose [89]. The GHb is retained on the columns, while the nonglycated hemoglobin passes through. In the next step the retained GHb is eluted from the column with a buffer containing a competing ligand such as sorbitol. Although it has been shown that a small proportion of GHb is not retained on the column [90] and the assay is sensitive to temperature change [91], the overall variation of replicate samples has been reported to be below 3%.

Method based on immunoassay

An immunoassay for HbA1c has recently become available [92]. The method, which is based on microtiter plate technology, utilizes a mouse monoclonal antibody raised against an epitope that consists of the ketoamine linkage plus the first eight amino acids of the N-terminal end of the hemoglobin β chain. The precision of this method is quite acceptable and the within-batch and between-batch coefficients of variation are equal to or less than 5%. One point to note is the low range of reference values (2.8–4.8%). The HbA1c values determined by the new immunoassay correlate well with other standard methods of HbA1c measurement, including HPLC. Since the antibodies used are specific for the underlying molecular structure of HbA1c, the immunoassay does not detect glycation of other hemoglobin variants, although samples should be treated to remove the labile component (Schiff base), since this can cause falsely elevated test results.
Frequency of HbA1c testing

HbA1c is thought to reflect average glycemia over the preceding few months [2] and has been demonstrated in multiple trials to be predictive of risk of diabetes complications [33,68]. Therefore, the ADA recommends that HbA1c testing be performed at least twice a year in patients who have stable glycemia and have met their glycemic goal, and quarterly in patients not at goal or in those who have had recent therapy changes. More frequent HbA1c testing may be necessary in patients who must be intensively managed, such as pregnant type 1 diabetic patients. Point-of-care HbA1c testing may be used for more timely monitoring [43].

HbA1c testing affected by comorbid illnesses

Hemoglobinopathies

There are a number of potential pitfalls in the use of HbA1c in the presence of a hemoglobinopathy or other complicating illnesses affecting the erythrocyte lifespan, such as hemolytic anemia (recently reviewed by [92,93]). In some clinical situations, commonly used clinical assay methods may yield false values (Tables 59.2 and 59.3). Understanding the methodology currently used by the practitioner’s clinical laboratory is essential to the proper interpretation of the test results. Over 700 hemoglobin variants are recognized worldwide; and some, like HbS, are common, affecting up to 9% of African Americans [94]. Since the presence of hemoglobinopathies can markedly affect HbA1c test results, close attention should be paid to the interpretation of laboratory results in this setting. Bry et al. [95] have recently reviewed the effect of hemoglobin variants on glycohemoglobin testing and provided practical guidelines for test applications and interpretation. If the HbA1c testing does not roughly correlate with the patient’s own monitoring of capillary blood glucose, the possibility of an underlying hemoglobin variant or disorder of erythrocyte stability should be explored. When dealing with populations in which HbSS, HbCC, or HbSC diseases are common and in which glycohemoglobin determinations have limited utility, laboratories should offer alternative forms of testing, such as glycated serum proteins or glycated serum albumin (discussed later). Physicians taking care of patients with diabetes should especially be aware of hemoglobinopathies that may be prevalent in their patient populations. A list of appropriate assays can be found at http://www.ngsp.org/interf.asp.

HbF comprises 70% of the total hemoglobin at birth and falls to less than 1% in adults. In some individuals with hereditary persistence of HbF, its concentration can reach up to 30% of total hemoglobin, while in patients with β-thalassemia and sickle-cell anemia, HbF can comprise from 2–20% of the hemoglobin level [94]. Most boronate affinity column techniques separate Hbf and GHB from HbA and HbA1c and provide accurate measurements of GHB; the immunoassay method and cation exchange, may show erroneous values [95]. Similarly, with HbS disease and trait, the boronate affinity method is the best assay to use (Table 59.3). In patients homozygous for sickle-cell anemia, interpretation of HbA1c requires caution, given the pathologic processes, including anemia, short red cell lifespan, and transfusion requirements. An alternative reference range may be used to correct for the shortened lifespan of red cells in HbS patients [96]. Additional measurements of glycomic control such as glycated serum albumin, should also be considered in these patients (see later).

HbC trait has a prevalence of 2.3% among African Americans and a false increase has been shown even with a boronate affinity method [95]. The use of an altered protocol on certain analyzers may fully separate HbC from HbA for accurate reading [97]. Calculation of mean blood glucose and use of fructosamine assay (see later) may be an alternative in these conditions to assess long-term control.

HbE trait is the most commonly encountered hemoglobin variant in Southeast Asia with a prevalence of up to 30% in the indigenous population [98]. Most of the cation exchange chromatographic methods will report falsely decreased values whereas boronate affinity methods and immunoassays are free of error in such cases. Again, a reduced erythrocyte lifespan and increased turnover of red blood cells needs to be considered in affected cases, which will impact the time span of the HbA1c estimate of mean glycemia.

Uremia

HbA1c levels have been found to be less reliable in renal failure because uremia per se has effects on HbA1c level independent of a change in blood glucose level. Urea present at high levels in patients with renal failure may dissociate in the circulation to form cyanate. Protonation of cyanate leads to formation of

---

Table 59.2 Conditions affecting measurement of GHB levels by chromatography

<table>
<thead>
<tr>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increase measured levels</td>
</tr>
<tr>
<td>– Negatively charged hemoglobin variants</td>
</tr>
<tr>
<td>– Uremia (carbamylation of Hb)</td>
</tr>
<tr>
<td>– Alcoholism</td>
</tr>
<tr>
<td>– Lead poisoning</td>
</tr>
<tr>
<td>– Elevated triglycerides</td>
</tr>
<tr>
<td>– Iron-deficiency anemia</td>
</tr>
<tr>
<td>– Post-splenectomy</td>
</tr>
<tr>
<td>– Hyperbilirubinemia</td>
</tr>
<tr>
<td>– Opiate addiction</td>
</tr>
<tr>
<td>– Chronic alcoholism</td>
</tr>
<tr>
<td>Decrease measured levels</td>
</tr>
<tr>
<td>– Positively charged hemoglobin variants</td>
</tr>
<tr>
<td>– Hemolytic anemias</td>
</tr>
<tr>
<td>– Acute or chronic blood loss</td>
</tr>
<tr>
<td>– Pregnancy</td>
</tr>
</tbody>
</table>

Source: Adapted from Schnedl et al. 2000 [94]; Eberentz-Lhomme et al. 1984 [101]; Service 1990 [105].
Glycated hemoglobin, serum proteins, and other markers as tools for monitoring

Table 59.3 Effects of frequently encountered hemoglobin variants on measurement of GHb

<table>
<thead>
<tr>
<th>Variant or derivative</th>
<th>Cation exchange</th>
<th>Immunoassay</th>
<th>Boronate affinity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbamyl-Hb</td>
<td>Increase typical</td>
<td>NC*</td>
<td>NC</td>
</tr>
<tr>
<td>HbAS</td>
<td>NC or increased</td>
<td>NC or increased</td>
<td>NC</td>
</tr>
<tr>
<td>HbAC</td>
<td>NC</td>
<td>NC or increased</td>
<td>NC or reduced</td>
</tr>
<tr>
<td>HbAE</td>
<td>Falsely low or new reference range needed</td>
<td>NC</td>
<td>NC</td>
</tr>
<tr>
<td>HbSS</td>
<td>Unsuitable</td>
<td>May be OK with new reference range</td>
<td>NC or new reference range needed</td>
</tr>
<tr>
<td>HbCC</td>
<td>Unsuitable</td>
<td>May be OK with new reference range</td>
<td>NC or new reference range needed</td>
</tr>
<tr>
<td>HbSC</td>
<td>Falsely low or unsuitable</td>
<td>NC</td>
<td>NC</td>
</tr>
<tr>
<td>Increased HbF</td>
<td>Falsely low or unsuitable</td>
<td>NC or false readings</td>
<td>NC</td>
</tr>
</tbody>
</table>

*NC = no change.

Source: Modified from Bry et al. 2001 [95]. Reproduced with permission of American Association for Clinical Chemistry.

isocyanic acid, which condenses with the N-terminal amino groups of hemoglobin in a process similar to glycation to form carbamylated hemoglobin [99]. Uremic patients may have carbamyl-Hb concentration as high as 3% of total Hb. Carbamylated hemoglobin tends to comigrate with HbA1c in assays separating HbA1c based on its electrical charge (cation-exchange chromatographic techniques) to give falsely high levels. This phenomenon only affects charge-dependent assays, such as column- and ion-exchange chromatography and agar gel electrophoresis, and the interference is only significant when serum urea levels are greater than 84 mg dL\(^{-1}\) (30 mmol L\(^{-1}\)) [100]. Immunoassay methods and boronate affinity columns to detect HbA1c are more reliable in uremic patients [95].

HbA1c levels also may be elevated in patients with renal failure because of an increase in rate of HbA1c formation, possibly as a result of uremic acidosis. Some analyzers, for example the A1c 2.2 HPLC analyzer, can apparently measure HbA1c accurately in uremic patients [101]. While HbA1c in the range of 6–7% may be estimated accurately in patients with severe renal impairment, values above 7.5% may overestimate the degree of hyperglycemia in patients with severe renal insufficiency [102]. These confounding factors arising from ion-exchange techniques can be avoided by use of boronate–agarose affinity chromatography or a thiobarbituric acid method, both of which can be used reliably in end-stage renal disease, but are time-consuming for routine determinations [103].

Overall, a boronate affinity chromatography method is the method of choice in the setting of hemoglobin variants and end-stage renal failure; however, affinity columns are unable to correct for the reduced lifespan of erythrocytes due to hemolysis or erythrocyte fragmentation. Clinicians also need to be aware of the changes in erythrocyte lifespan occurring during rapid blood loss or initial treatment with iron, folic acid, or vitamin B12 of the respective deficiency anemias [104].

Other metabolic conditions
The presence of an elevated bilirubin or hyperlipidemia interferes with certain HbA1c assays, especially in situations when whole blood is used rather than an erythrocyte hemolysate [26]. Marked hypertriglyceridemia, hyperbilirubinemia, opiate addiction, lead poisoning, alcoholism, and long-term treatment with large doses of aspirin may spuriously elevate HbA1c measurements when determined by cation-exchange chromatography [105]. It has also been shown that vitamin C in doses of 1.0 gram daily for 3 months decreases HbA1c as measured by affinity chromatography by about 20%, while it increased HbA1 determined by agar-gel electrophoresis by 16% [106,107].

Glycation of serum proteins
The measurement of blood glucose is the *sine qua non* for the detection, treatment, and monitoring of the progression of diabetes. The discovery of nonenzymatic glycation of hemoglobin, discussed earlier, paved the way for the possibility that other proteins could also be derivatized by glucose and serve a possible role as a measure of glycemia over time. Similar to hemoglobin, serum and intracellular protein molecules are subject to nonenzymatic glycation to form stable ketoamines, or fructosamines [108]. Fructosamine is the standard name for 1-amino-1-deoxy-fructose, also called isoglucosamine by Emil Fischer, who first synthesized the compound in 1886. The glycation of albumin accounts for approximately 90% of glycated serum proteins [109]. Glycation of serum proteins is similar to that of hemoglobin, except serum proteins do not contain β-amine terminals; therefore, attachment of glucose occurs at the ε-amino groups of lysine side chains [110]. Consequently, charge-dependent methods, like those used to separate HbA1c from HbA, cannot be used to measure fructosamine.
The rate of stable ketoamine formation is four to five times faster in glycate serum proteins than in hemoglobin [111]. Fructosamine reflects integrated glycemia during a period of 2–3 weeks, reflecting the half-lives of the derivatized serum proteins [112]. Use of the fructosamine assay has been well documented in clinical settings like hemolytic anemias and hemoglobinopathies where measurement of HbA1c by conventional techniques may not provide a reliable index of glucose assessment (discussed earlier). For this reason, measurement of fructosamine has been on the rise especially in the developing world where significant numbers of patients have anemia and abnormal hemoglobin. In addition, fructosamine testing may be very helpful in patients with chronic renal failure who have increased levels of carbamylated hemoglobin, which can affect the clinical HbA1c measurement assay [93]. However, falsely low fructosamine values in relation to mean blood glucose values will occur with rapid albumin turnover as occurs in patients with protein-losing enteropathy or nephrotic syndrome [93].

Assay methods for glycated serum proteins
In 1982, the fructosamine assay was developed and became commercially available [113]. To date, reduction of nitroblue tetrazolium dye (NBT) has been the most common method to estimate glycated serum proteins (fructosamine). It is a colorimetric test that is based on the ability of fructosamines to act as reducing agents in alkaline solution and convert NBT to a purple dye that can be monitored by absorbance at 530 nm. As with any other biochemical method, fructosamine assays have their limitations [114]. A number of problems have been found with this original automated fructosamine assay, such as standardization difficulties [115], lack of specificity [116], and interference by lipemia and urate [117,118]. Of course, fructosamine values are affected by low serum albumin and plasma proteins, as in cases of protein-losing nephropathy or liver failure, and must be interpreted with caution under these circumstances.

A variety of biochemical methods have been developed for fructosamine [119,120]. In the early 1990s, Boehringer–Mannheim and Roche Diagnostics developed an improved fructosamine assay [121], with reduced protein matrix effects and less interference by lipemia and urate. More recently, a nonseparation fluorescence-quenching assay for the measurement of total glycated serum proteins has been introduced [122]. The principle of this method is that the fluorescence of a fluorophore, a boronic acid derivative, is quenched when it is bound to glycated proteins but not when it is unbound. The drop in fluorescence is proportional to the concentration of glycated protein.

Clinical usefulness of fructosamine analysis in diabetes mellitus
Fructosamine assays may be used in clinical settings to detect short- to medium-term changes in mean blood glucose. Fructosamine assays also assist in certain clinical conditions where measurement of HbA1c is considered less reliable. The half-life of hemoglobin is approximately 60 days, of albumin 14–20 days, and that of various other proteins, 2.5–23 days. In a group of diabetic patients treated to achieve better glycemic control, the concentration of glycated albumin entered the normal range and plateaued, while the concentration of HbA1c continued to fall [123]. This study showed that the kinetics of the fructosamine test differ from those of the HbA1c, and that it might serve as a more sensitive indicator of average blood glucose over a shorter time frame.

Austin et al. [124] showed that measurement of fructosamine could be a useful adjunctive test for management of outpatients with diabetes in situations where it is not practical to obtain same-day glycohemoglobin concentrations. This study also proved useful for patients less compliant with home glucose monitoring, or where recent changes in insulin dose or clinical presentation might not be reflected in the HbA1c levels. Weekly home testing of fructosamine, combined with therapeutic interventions based on the results, has also been found to lead to a more rapid and significant improvement in glycemic control than did a conventional regimen of testing based solely on capillary glucose testing [125].

The correlation between predicted HbA1c and fructosamine was fair ($r = 0.88$) in diabetic patients but not in control subjects ($r = 0.01$) [126]. Other studies have questioned whether there is any additional benefit gained from measuring the fructosamine in the course of clinical monitoring of patients with diabetes that cannot be obtained from measuring capillary glucose or HbA1c [127,128]. A recent study has substantiated that HbA1c measurement correlates better with home capillary blood glucose levels than with the fructosamine assay, even over the previous 2–3 weeks [125]. The most commonly cited problem confounding the interpretation and reliability of such data is related to short-term fluctuations in serum protein levels observed in diabetic patients [129,130]. Thus, although tests such as the fructosamine assay are relatively inexpensive and simple to perform, they have been questioned as a reliable index of hyperglycemia. Accordingly, fructosamine is not sufficiently sensitive or specific to be used as a screening tool for the diagnosis of diabetes. Although similar conceptually, fructosamine and GHB are two very distinct tests with different reference ranges ($170–285 \mu\text{mol}\cdot\text{L}^{-1}$ in a nondiabetic patient, and 0.6–3.0%, respectively) [131,132].

Monitoring of glyceremia in pregnancy

Measurements of HbA1c are recommended with greater frequency during pregnancy with preexisting diabetes. The ADA specifically recommends a DCCT-referenced assay performed at the initial visit, monthly throughout pregnancy until target levels of <6.0% are achieved, and then every 2–3 months [133]. However, although subject to the potential confounding influences of variations in serum protein concentration,
glycated serum protein measurements may be conceptually more appropriate than HbA1c given the time constraints of normal gestation and the desire to maintain near-normal glycemia over short-term periods. Because of changes in the distribution of body water, a significant decrease in serum albumin and serum fructosamine concentration occurs during normal pregnancy [134]. In spite of this, the value of short-term monitoring of glycemic control has been substantiated in a few studies [135,136]. In general, the use of glycated serum protein assays in diabetic pregnancy cannot be used to replace the frequent monitoring of capillary glucose levels, but perhaps as a verification of the overall level of glycemic control that may lead to alterations in the treatment plan if the glycated proteins begin to increase with monitoring every 3–4 weeks [137].

**1,5-Anhydroglucitol as a measure of glycemia**

1,5-Anhydroglucitol (AHG) is a polyol mostly derived from ingestion that is competitively excreted with glucose by the kidneys. As glucose levels rise, the amount of AHG decreases. After extensive studies and use in Japan and subsequent studies in the US, AHG testing was recently approved by the FDA for clinical use (GlycoMark™). AHG is emerging as a marker of glycemia that responds more readily to glucose trends than HbA1c, and also offers the advantages of a simple blood test that is metabolically stable and shows small biological variability [138]. AHG may be a more valid predictor of the risk of diabetes complications than fructosamine, but this remains to be proven. Also, available US studies have excluded patients with renal disease, and AHG may be unreliable in these patients. US studies show that serum AHG reflects glucose levels over the prior 48 hours to two weeks, and demonstrates good concordance with HbA1c and fructosamine levels and agreement with 20 years of previous Japanese studies [138–140]. Changes in AHG with changes in glucose are also more significant than with fructosamine in these studies. McGill et al. propose AHG as an appropriate interim marker for use between 3-month HbA1c measurements to help patients modify behaviors before complications occur [138]. AHG may also be a good indicator of postprandial hyperglycemic excursions. Finally, Sone et al. advocate for the use of AHG to identify patients at risk for future uncontrolled hyperglycemia and diabetic complications [141].

**Advanced glycosylation end-products and diabetic complications**

Multiple mechanisms have been described by which chronic hyperglycemia may contribute to the pathological end-organ complications that occur in diabetes mellitus. These include direct effects of elevated glucose on cells, hyperosmolality, oxidant stress, and nonenzymatic glycosylation. The irreversible ketoamine linkage that follows Schiff base formation of glycated proteins can undergo a series of multistep reactions potentially along several different pathways (Figure 59.3). Some of the formed products involve fragmentation to form deoxyglucosones, which may then lead to additional protein cross-links [143,144]. The deoxyglucosones can form highly reactive dicarbonyl compounds that cyclize to form electrophilic pyrrole intermediates and more advanced cross-linked fluorescent derivatives [4,93,145]. The complex biochemistry of the AGEs and the implication of the formation of these derivatives in vivo is incompletely understood, and remains an area of active research.

Once formed, AGE-protein adducts, such as carboxymethyllysine (CML) and pentosidine are biochemically irreversible and remain quite stable in the body. AGEs are a heterogeneous family of biochemical end-products of nonenzymatic glycosylation. Circulating human AGE peptides are of variable size (molecular weight between 2000 and 6000) and characterized by intra- or intermolecular cross-linking products with cyclic aromatic rings that convey fluorescence [93]. In addition to their chemical reactivity, some AGE structures, for example N-carboxymethyllysine [146], lead to the formation of antibodies in vivo [147], which may portend an additional mechanism for tissue damage. AGE modification of proteins is potentially involved in the normal aging process as well as in the pathogenesis of diabetic nephropathy, atherosclerosis, diabetic retinopathy, and chronic renal failure [148–152]. An additional element of complexity in the effects of AGE products has been the discovery of a specific AGE receptor, called “RAGE,” on a variety of cell types including macrophage/monocytes [153] and endothelial cells [154]. The AGE receptor may transmit effects of AGE products on signal transduction that may mediate its detrimental effects in tissues [145–156].

The discovery of RAGE has led to the development of a whole-cell radioreceptor assay for AGEs [157]. The principle of this assay rests on the assumption that the family of AGEs have similar affinity for the AGE receptor, and that those present in the sample will compete, in proportion to their concentration, with the binding of labeled AGE-modified albumin. The amount of AGEs in the sample is read from the standard curve showing competition binding of labeled and unlabeled AGE-modified albumin, and expressed in units defined as amount of the test protein required to inhibit the binding of 50% of the labeled AGE-modified albumin [157]. With this assay it is possible to detect AGEs in serum both in normal and diabetic patients. In the first report of its kind it has been documented that serum AGEs of molecular mass <10kDa in nondiabetic patients with end-stage renal disease are increased 1.5-fold compared to those in normal individuals, and increased fivefold in patients with end-stage renal disease and diabetes [158]. This study raises two important questions, and the answers remain uncertain. First, do the AGEs themselves contribute to the development of chronic complications seen in diabetes? The authors of the cited
Figure 59.3 Formation of advanced glycosylation end-products from glucose. N-glucosylamine is a reversible Schiff base product and the Amadori rearrangement leads to the formation of 1-amino-1-deoxyketose. The production of these reversible early products can give rise to irreversible advanced products through generation of highly reactive carbonyl compounds such as 3-deoxy-D-gluconone. Source: Brownlee 1992 [142]. Reproduced with permission of the American Diabetes Association.

report suggest that the AGEs in the serum (especially those of low molecular weight) are metabolites of AGE-modified proteins formed in the tissues. The major route of elimination of these metabolites, especially those of molecular weight of <10 kDa, is through the kidneys, and the progression of renal failure leads to their retention in serum. The several-fold increased retention of low-molecular-weight AGEs in patients with end-stage renal disease and diabetes compared to the other study groups indicates that the turnover of AGEs in diabetes is greatly increased. Moreover as shown by the authors [158], the serum AGEs are poorly dialyzable and may contribute to the rapid progression of complications and mortality of patients with diabetes requiring dialysis.

At present, monitoring of serum AGEs remains a research tool that is not widely available. In the next few years, progress in this area of research may provide a new dimension in the clinical monitoring of diabetes and its complications that can extend the value of measurement of glycated proteins far beyond their present applications. A number of potentially therapeutic compounds are under extensive study in animals and humans for potential use in diabetics, in an effort to eliminate the accumulation of AGE products or minimize the deleterious effects of AGE [93]. Current early-stage research indicates that inhibition of AGE formation and receptors for AGE (RAGE) may represent possible therapeutic endpoints for new therapies to treat diabetic retinopathy [159].

Artificial pancreas

The advent of the continuous glucose monitoring (CGM) device, which measures glucose in the interstitial fluid at least 288 times daily, has made possible the concept of the “artificial pancreas” when linked with previously developed insulin pump technology for continuous subcutaneous insulin infusion (CSII) (or in a bihormonal system with insulin pump plus glucagon pump for counterregulation) that is controlled by a computer-run insulin delivery algorithm. The premise of the artificial pancreas is the use of a closed-loop system to eliminate SMBG, monitoring with surrogate markers such as HbA1c, and self-adjustment of insulin [160], while concurrently helping diabetic patients reach previously elusive near-normal HbA1c levels with fewer extremes of hypo- and hyperglycemia and diabetic complications [161]. In other words, the artificial pancreas would attempt to replicate the functioning of the β-cells of the pancreas of a person without diabetes. Additionally, the artificial pancreas would obviate the need for available pancreas or islet cells for transplantation, immunosuppression for those transplantations, and dependence on the long-term functioning of a transplant [161]. Two main approaches to the closed-loop system have been described in the literature: extracorporeal (with both subcutaneous glucose monitoring and insulin delivery) and implantable (with intravenous sampling and intraperitoneal insulin delivery). Both of these approaches have been shown to be attainable in the laboratory setting and with use of a model predictive controller [162]. Limitations to use in the home setting include reliability and accuracy of CGMs, unwieldy implanted systems and need for continuous intravenous access, regulatory difficulties, and disagreement over extracorporeal versus implantable approach as well as associated insulin delivery calculation algorithms [162]. A particular challenge involves estimation of pre-meal boluses without information as to meal size or timing in a fully closed-loop system. Revisions involve inclusion of meal size and meal timing information in a semi-closed-loop approach, or timing but not size of the meal in a so-called “closed-loop with qualitative meal announcement approach” [162], with most closed-loop systems appropriating one of the first two approaches.

Current studies have utilized in silico design for regulatory approval via the FDA, using computer-based simulations on virtual subjects as a substitute for animal trials, allowing for simulation of any open- or closed-loop scenario prior to testing in human subjects [163]. Multiple studies have shown that the amount of time that BGs remain in the prescribed
target ranges increases with closed-loop delivery in human subjects [163,164]. Model-predictive control with utilization of in silico design may soon allow for the institution of a truly viable closed-loop system without need for patient-based SMBG or monitoring of serum markers for control of blood glucose.

Acknowledgments

The authors are grateful to Drs. Jerzy Kolaczynski and Barry Goldstein for their contributions to the writing of previous versions of this chapter.

References


The Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications (DCCT/EDIC) Study Research Group: Intensive diabetes treatment and cardiovascular


Glycated hemoglobin, serum proteins, and other markers as tools for monitoring

SECTION XI

Complications of diabetes: general and microvasular
Pathogenesis of diabetic microvascular complications

Josephine M. Forbes1 and Brooke Harcourt2
1 Cardiovascular and Metabolic Diseases Mater Research Institute, The University of Queensland, Brisbane, QLD, Australia
2 Mater Research Institute - UQ, TRI, South Brisbane, QLD, Australia; Mater Clinical School, University of Queensland, Brisbane, QLD, Australia

Introduction

The development of chronic vascular complications is common in individuals with both type 1 (T1) and type 2 diabetes (T2DM). These are in general, divided into micro- and macrovascular complications, with the most prevalent microvascular complications being kidney disease, blindness, and amputations. Given the morbidity and mortality associated with these complications, there is a significant treatment gap, with current therapies only slowing disease progression. In addition, a number of large-scale clinical trials have in general yielded disappointing or incremental results with respect to the benefits of current treatments. One fundamental observation, however, in clinical studies, is that early strict glycemic control can radically reduce the incidence of microvascular complications and that this is further enhanced by addressing abnormal lipid profiles and hypertension. The benefits of improved glycemic control are less consistent with respect to macrovascular disease, where large-scale clinical trials in patients with a long duration of diabetes have implied that strict glycemic control may in fact be detrimental in some circumstances. Indeed this suggests that there may be two distinct phases in the development of diabetic complications.

Therefore given the clear association between strict glycemic control and microvascular disease in diabetes, it seems that more careful consideration of the consequences of poor glycemic control on the microvasculature of the kidney, nerves, and retina is warranted. As a consequence, pathologic pathways uniquely activated by hyperglycemia at sites of diabetic complications [1] and differentiating these from tissue/organs sites which appear less susceptible to hyperglycemia. Each of these concepts will be discussed within this chapter as we outline common pathogenic pathways (Figure 60.1) which may play causative roles in diabetic microvascular complications.

Key points

- Glycemic control is a major determinant of progressive microvascular complications.
- Microvascular complications and glycemic control should not be viewed as mutually exclusive when considering therapeutic options.
- Despite different clinical manifestations probably common pathogenesis of disease at unique sites of microvascular disease.
- Common accelerators include blood pressure, abnormal lipid profiles, obesity, gender, age.
- Logical to study patterns seen in all microvascular complications, rather than one disorder in isolation.
- Pathogenic pathways directly affected by altered glycemic control—most logical therapeutic targets.
- Should consider outcomes of therapies for all microvascular complications when considering their effectiveness.
the EDIC study (Chapter 41) on microvascular endpoints. Of strict glycemic control persisted for another 10 years in patients were followed (described in Chapter 41) the benefits of the DCCT where both strictly and conventionally controlled changes from strict to more conventional glycemic control. In well beyond study completion and the time when there were diabetic microvascular complications. These effects are evident has a sustained impact on reducing the risk of subsequent large clinical trials where early intensive glycemic control
glycemia. Metabolic memory has been observed in a number episodes of either good or poor glycemic control. The most likely explanation for this phenomenon, termed “metabolic memory,” is an epigenetic mechanism involving glucose-induced effects on histone modifications leading to modulation of vascular gene expression [3,4] which persists despite a return to normoglycemia. Metabolic memory has been observed in a number of large clinical trials where early intensive glycemic control has a sustained impact on reducing the risk of subsequent diabetic microvascular complications. These effects are evident well beyond study completion and the time when there were changes from strict to more conventional glycemic control. In the DCCT where both strictly and conventionally controlled patients were followed (described in Chapter 41) the benefits of strict glycemic control persisted for another 10 years in the EDIC study (Chapter 41) on microvascular endpoints. Metabolic memory has also been described in cultured cells [3] and in animal models showing that islet transplantation cannot prevent retinopathy progression beyond a certain time point [2,5,6]. It is still unresolved as to what is the specific cause of metabolic memory. One promising candidate is the programming of a reversible epigenetic memory effect on cells via good glycemic control.

Gene regulation

Can our tissues remember?
The contribution of hyperglycemia per se to the complications of diabetes needs to be further investigated given the relatively disappointing findings from the ACCORD and ADVANCE clinical trials, which explored the effects of strict glycemic control in type 2 diabetic subjects with established cardiovascular disease (described in Chapter 42). On one hand, it could be that a lack of effect on cardiovascular mortality relates to the relatively short duration of the trials, but equally plausible is that irreversible vascular changes have occurred prior to the trial, such as may be imparted by biochemical reactions, for example advanced glycation. The first suggestion of this is from some elegant studies performed in diabetic dogs with retinopathy more than 20 years ago [2], which showed progression of retinal disease despite good glycemic control.

This is now postulated to be the result of hyperglycemic memory or a legacy effect where the body remembers previous episodes of either good or poor glycemic control. The most likely explanation for this phenomenon, termed “metabolic memory,” is an epigenetic mechanism involving glucose-induced effects on histone modifications leading to modulation of vascular gene expression [3,4] which persists despite a return to normoglycemia. Metabolic memory has been observed in a number of large clinical trials where early intensive glycemic control has a sustained impact on reducing the risk of subsequent diabetic microvascular complications. These effects are evident well beyond study completion and the time when there were changes from strict to more conventional glycemic control. In the DCCT where both strictly and conventionally controlled patients were followed (described in Chapter 41) the benefits of strict glycemic control persisted for another 10 years in the EDIC study (Chapter 41) on microvascular endpoints. Metabolic memory has also been described in cultured cells [3] and in animal models showing that islet transplantation cannot prevent retinopathy progression beyond a certain time point [2,5,6]. It is still unresolved as to what is the specific cause of metabolic memory. One promising candidate is the programming of a reversible epigenetic memory effect on cells via good glycemic control.

Histone modifications

There are two major processes which contribute to the remodeling of chromatin, a complex of DNA and histone proteins. First, posttranslational modifications of the histone tails via acetylation, methylation, advanced glycation, ubiquitylation, and phosphorylation can occur in diabetes at sites of microvascular disease [7,8]. Second, DNA can be directly altered via modulating methyl groups within CpG sites. Ultimately, the role of these processes is to allow or to prevent the unwinding of DNA to manipulate the exposure of specific gene sequences. These effects may be heritable and thus are able to be passed through many cell divisions.

Acetylation of histones is a dynamic process regulated by two opposing families of enzymes, histone acetyltransferases (HATs), which are responsible for histone lysine acetylation resulting in “opening” of the DNA to allow for transcription factor and RNA polymerase II binding, while histone deacetylases (HDACs) remove lysine acetylation and repress gene transcription [7,8]. Indeed, glucose concentrations are known to alter histone acetylation.

Methylation of histones at lysine and arginine residues, although also a dynamic process, is thought to impart longer lasting changes which lead to both repression and activation of gene transcription. Protein arginine methyltransferases (PRMTs), are commonly involved in gene regulatory events via mono- or dimethylation of arginine residues [7,8]. At lysine residues, methylation is often complex given that they can be mono-, di-(me2), or trimethylated. Lysine demethylases (HDMs) can also alter gene expression emphasizing the bidirectional nature of histone methylation [7,8]. DNA methylation also changes in response to transient hyperglycemia.

A number of changes in expression and activity of histone deacetylases (HDAC), histone methyl transferases (HMTase), histone acetyl transferases (HATs), and histone demethylases (HDMs) have been identified at sites of diabetic complications. Most of these studies have been performed in endothelial cells, where changes in these enzymes have been associated with the regulation and transcription of specific genes including the NF-κB subunit, p65 [3]. Preclinical studies in leukocytes including monocytes from diabetic patients have exhibited epigenetic modifications including changes in effects on histones such as H3K9me2 and H3K4me2, which are associated with immune and inflammatory pathways [7,8]. In addition, TGF-β1 treatment of renal mesangial cells increases the histone methyltransferase, SET7/9, which is associated with the expression of profibrotic genes in these cells [7,8].
Reversal of epigenetic memory, or epigenetic therapy, has gained interest for the treatment of diabetic complications. Indeed, curcumin (a derivative of turmeric), which is an inhibitor of histone acetyl transferases, can lower the expression of a number of inflammatory genes and has shown promise in the treatment of diabetic nephropathy in both humans and in experimental models of diabetes [1].

Other approaches targeting epigenetic regulation in experimental models of diabetes [7,8] have also shown promise. Furthermore, it is likely that these compounds may be useful for the treatment and prevention of atherosclerosis and retinal neovascularization given previous beneficial effects seen in the nondiabetic context with respect to these disorders [1].

**MicroRNA**

MicroRNAs (miRNAs) are short sequences of RNA, which can manipulate the translation of messenger RNA via interaction with complementary sequences located within the 3′ untranslated region. MicroRNAs are transcribed from sequences contained within the “noncoding” introns of DNA and play important roles in regulating gene expression.

The diabetic kidney has received the most scrutiny with respect to miRNA expression and effects. Not surprisingly, most of this research has focused on miRNAs that target molecules involved in renal fibrosis via effects elicited by TGF-β1 such as miR-192 which is thought to target and repress zinc finger E-box-binding homeobox 2 (ZEB2), affecting changes in deposition of type 1 collagen, most likely via repression of Smad7 signaling [9]. This has also been shown in other renal diseases. A number of other miRNAs are also implicated in the accumulation of extracellular matrix in organs effected by diabetes, including miR-21, miR-29, miRNA-216a, miR-377, and miRNA-93 via regulation of a number of TGF-β stimulated molecules [9]. In addition, protection against the accumulation of collagen I and fibronectin and decreases in VEGF expression have also been shown in the diabetic kidney and retina by the miRNA 200 family [1,10,11].

**Hyperglycemia—too much fuel or not enough?**

The chronological sequence of events which ultimately results in a decrease in cellular energy production, despite excess fuel (glucose and lipids) availability, is not well understood in diabetes. It is widely postulated that cells within tissues that are prone to diabetic complications have difficulties modulating glucose transportation and metabolism. Initially, changes in the diabetic milieu alter energy production in these cells by increasing concentrations of intracellular glucose-derived metabolites. Eventually, this progresses to impairment of cellular energy production by inhibition of energy substrate delivery to cells because of cellular compensatory mechanisms such as insulin resistance. However, this ultimate loss of ATP content within complication-prone tissues is considered to occur relatively late in the development of the disease and postulated to be responsible for the end-organ dysfunction and cell death [12]. Conversely, however, it is plausible that changes which actually enhance energy production from excesses in substrates such as glucose in particular, are more likely early pathologic events in the development of complications. It is possible to hypothesize therefore, that modulating glucose delivery into tissues thereby preventing the switching of cells to inappropriate fuel sources is likely to be another reason as to why glycemic control appears effective in lowering the risk of microvascular complications.

Unfortunately, optimization of glycemic control in T2DM requires therapeutically complex regimens where outcomes of individual agents are often difficult to assess. It has been shown that at least some of the antiglycemic agents exhibit direct benefits on microvascular complications (Figure 60.2). This includes thiazolidinediones, agonists for peroxisome proliferator activated γ receptors (PPARγ agonists) have shown beneficial effects on complications, independent of their glucose lowering through attenuating fibrosis by reducing the production of pro-fibrotic cytokines. The future of this class of agent in diabetes, however, has been overshadowed by data showing increased risk of cardiovascular events with rosiglitazone (see Chapter 40) and increased bladder cancer with pioglitazone as well as other side effects with both agents such as fluid retention, exacerbation of heart failure, and increased bone fractures.

Metformin is more complex, where conflicting effects on the microvascular complications of diabetes, have been identified. As an example, metformin therapy worsens peripheral neuropathy in diabetic individuals, which is postulated to be the result of effects on vitamin B12 (see Chapter 41). It is interesting that detrimental effects of modulating specific vitamin B groups have also been shown in diabetic nephropathy in humans (see Chapter 41). Conversely, other studies have identified beneficial effects on diabetic renal disease with metformin in humans (Chapter 41) and in animal models [1]. In general, these favorable outcomes are thought possibly to be as a result of effects on lipids including free fatty acids (FFA) leading to a less pro-inflammatory profile and reduced oxidative stress.

Analogs that mimic the actions of the insulinitrophic glucagon-like peptide (GLP-1) have been shown to bind to receptors present at sites of diabetic complications [13]. In addition, activation of the GLP-1 receptor using these GLP-1 receptor agonists in diabetes appears to be beneficial for kidney disease, and appears to be mostly independent of changes in glucose homeostasis. GLP-1 receptor agonists also affect sodium handling by the kidney which is another important contributor to development and progression of diabetic complications by modulation of hemodynamic pathways [13].

Another way to increase the amount of GLP-1 within the circulation is to target its metabolizing enzyme, dipeptidyl peptidase IV (DPP4IV). DPP4IV inhibitors have also been shown to have benefits on diabetic renal disease [13].
Glucose metabolism

Glucokinase is part of the hexokinase family and is an important rate-limiting enzyme involved in the transport of glucose into cells across the plasma membrane. Changes in this rate-limiting enzyme have been observed at sites of diabetic complications. Once inside the cell, the first major energy system into which glucose-derived molecules are channeled is glycolysis. Glycolysis is an anaerobic process and there is evidence that this pathway is disrupted in diabetes [12]. Glycolysis, however, is not an efficient way to derive energy from glucose, with only four adenosine triphosphate (ATP) molecules being generated for each molecule of glucose. Hence, our cells shuttle the pyruvate produced from glycolysis into our cellular power stations, the mitochondria, where it is used for oxidative phosphorylation by the mitochondrial respiratory chain generating up to 15 times more ATP per molecule of glucose. FFA and ketoacids, which enter the cell, are also taken into the mitochondria and used for oxidative phosphorylation following beta-oxidation. However, glucose-derived intermediates are the most efficient facilitators of ATP generation and use less oxygen than substances such as FFA and ketoacids and hence are the most common fuels used in cells prone to microvascular complications. Conversely, studies such as ADVANCE, VADT, and ACCORD (see Chapter 42), have shown an apparent lack of effect of glycemic control in preventing the progression of CVD and one could postulate that this is as a result of the high utilization of FFA at sites of macrovascular complications, although this remains to be determined.

An electrochemical proton gradient is created from energy released from nutrients as electrons, which flow through the mitochondrial respiratory chain. In some circumstances, rather than generating ATP, this group of reactions can be uncoupled, producing heat or stopping ATP production. It is therefore not difficult to imagine how this tightly controlled group of reactions may be affected by diabetes where there is excess substrate availability and profound changes in cellular behavior. It remains to be determined, however, as to why some cells seem more vulnerable to these changes, resulting in microvascular changes that we see characteristically in diabetes. Indeed, changes in energy production are thought to be major contributors to the development of diabetic microvascular complications [12]. Some examples shown to date include altered delivery of substrates, switching of cell-specific fuel sources among glucose intermediates, fatty acids and ketoacids, altered mitochondrial respiratory chain function and uncoupling of the respiratory chain. Dysregulation of the family of proteins that regulate uncoupling has also been previously reported at sites of diabetic complications [12].

Other cellular pathways for glucose metabolism

Once glucose is transported inside the cell, most of it is metabolized via glycolysis, through steps involving the conversion of glucose to pyruvate, and then to acetyl-CoA, which enters the citric acid cycle for further energy production through oxidative phosphorylation in the mitochondria.
of glucose-6-phosphate to fructose-6-phosphate. When intracellular glucose concentrations are high, however, glycolysis can divert fructose-6-phosphate step-wise to UDP (uridine diphosphate) N-acetyl glucosamine. N-acetylglucosamine is used for posttranslational modification of proteins within the cytosol and nucleus by single O-linked N-acetylglucosamine (O-GlcNAc) glycosylation (which is discussed in detail later in protein trafficking). These resulting sugar residues can compete with phosphate groups altering gene expression in diabetic tissues [1].

Aldose reductase

Under normal glucose homeostasis, aldose reductase has a physiologic role in detoxification of aldehydes into inert alcohols. Hyperglycemia was shown to increase cellular flux through the sorbitol/polyol pathway more than 40 years ago [14]. Under these conditions, NAD(P)H, delivered from the pentose phosphate pathway, catalyzes the conversion of glucose to sorbitol by aldose reductase rather than altering the flux of glucose into glycolytic pathways. Facilitation of this alternate pathway of glucose flux is thought to contribute to cellular antioxidant inhibition by depleting reduced glutathione and glutathione peroxidase enzymes. Changing cellular concentrations of sorbitol may also provide a source of excess NADH to complex I of the mitochondrial respiratory chain. In addition, intracellular accumulation of sorbitol alters cellular osmolarity exacerbating cellular oxidative stress. In support of this, there is some evidence that diuretics which decrease osmotic stress can protect from cell death in cells exposed to hyperglycemia.

In animal models of diabetic microvascular complications, there has been some success achieved in preventing disease progression using inhibitors of aldose reductase [15,16]. However, this has failed to translate to the clinical context despite decades of active investigation (see Chapter 41).

Impaired insulin signaling

It is widely thought that exogenous insulin therapy as part of more intensive glycemic control can protect against the development and progression of microvascular complications in diabetic individuals (see Chapter 41), although this must be weighed against what appear to be increases in certain cardiovascular risk factors such as weight gain (Chapter 42). What is noteworthy, however, is that in experimental models, intranasal insulin [17] and pancreatic islet transplants show superior protection to other methods of insulin administration. This implies that other factors related to endogenous insulin secretion and trafficking such as C-peptide, glucagon and immune modulation (discussed later) may have important roles in the protection against microvascular complications. This is in line with recent data from the “Joslin 50-year medalist” cohort which also suggests that remnant pancreatic β cells may protect against the development of microvascular complications (Chapter 41).

The cellular loss of signaling in response to the hormone insulin is termed insulin resistance. Despite a large focus on insulin resistance at sites of peripheral glucose uptake such as skeletal muscle, the liver, and adipose tissue, there are many cell types that require insulin-mediated glucose uptake for normal function including sites of microvascular complications. One study has demonstrated that a deficiency in insulin receptor signaling in podocytes of the kidney can induce a disease state reminiscent of diabetic nephropathy even in the setting of normoglycemia [18]. Selective targeting of glucose transporters such as SGLT2 and GLUT-1 in nephropathy has also shown some benefit in experimental models of diabetic renal disease [1].

Experimentally, deficiencies in insulin signaling can be assessed by examining a reduction in serine 473 Akt phosphorylation in insulin-target tissues. It is interesting that a loss of serine 473 Akt phosphorylation has been identified at sites of diabetic complications [18,19] particularly in models of T2DM. Conversely, Akt activity is increased in some tissues and vascular beds affected by complications in T1DM. Whilst it is likely that impairment of Akt activity is preventable with strict metabolic control, the combination of hyperglycemia and insulin treatment results in enhancement of mTOR activity which may also be of benefit (discussed in the Autophagy section later). An adequate explanation for this paradox at sites of diabetic complications remains to be established but may be resolved by evaluation of therapeutic benefits of pharmacologic modulators of Akt activity.

Hemodynamic pathways

We know that there are many factors that contribute to the development and progression of diabetic microvascular complications, including metabolic and hemodynamic factors [1]. The best characterized of these hemodynamic contributors include systemic and tissue-derived components of the renin-angiotensin-aldosterone system (RAAS) which are actively targeted as a major part of clinical care for diabetic individuals who develop microalbuminuria or hypertension (Chapter 41). Although hypertension is often considered a manifestation of diabetic nephropathy, it is also an important systemic factor in exacerbating or promoting diabetic microvascular complications.

The renin-angiotensin-aldosterone system (RAAS)

The RAAS is a hormonal cascade that is considered as a body-wide master controller of blood pressure and fluid balance. In concert with these systemic effects, a local tissue RAAS appears to be present at sites of microvascular complications. Clinically, the most widely applied RAAS blockers for microvascular disease in diabetes interrupt the conversion of angiotensin I to angiotensin II (AngII), namely angiotensin-converting enzyme-1 inhibitors (ACE-I), agents
which compete with Ang II for binding to the AT1 receptor, AT1 antagonists (ARB), or drugs that inhibit binding of aldosterone to the mineralocorticoid receptor (aldosterone antagonists). The clinical trials outlining the evidence which justifies the use of these agents in humans are discussed in Chapter 41. In the following, we outline some of the newer aspects of the RAAS which have shown promise in experimental models in combating diabetic complications.

Preprorenin, the active precursor of renin, is synthesized and stored within the juxtaglomerular (JG) apparatus of the kidney cortex. Upon stimulation from sources such as baroreceptors and other endocrine pathways, prorenin is secreted through fenestrated capillaries which elevate renin concentrations within the circulation. Another arm has recently been shown experimentally to contribute toward renin activation and secretion under pathologic conditions in experimental models. This pathway is mediated through changes in succinate uptake and utilization via the GPR91 receptor in the collecting ducts which are located in close proximity to the JG apparatus of the kidney [20]. These studies suggest that targeting GPR91 is beneficial in diabetic nephropathy in addition to the finding that urinary succinate concentrations may be a potential biomarker of disease development.

Renin is the hormone responsible for the cleavage of angiotensinogen to angiotensin I (AngI) and this reaction occurs primarily in the systemic circulation. There are a number of pathways that are thought to influence the secretion of prorenin. These include renal pressure sensors (baroreceptors), endocrine pathways, and intracellular mechanisms which may be independent of systemic blood pressure. It appears that the renin inhibitor aliskiren does have some benefits; however, there is some concern as to the side effects of these agents when they are used in combination with ACE inhibitors or angiotensin receptor blockade. Antagonists of the (pro)renin receptor have also been shown to be renoprotective in animal models.

Most commonly, AngI is cleaved to AngII by angiotensin-converting enzyme-1 (ACE-1), where AngII is arguably the most potent effector hormone of the RAAS exerting direct effects on mean arterial blood pressure and extracellular fluid volume. ACE-1 is localized at a number of sites of diabetic complications and may be regulated renally and hepatically via midkine. Angiotensinogen can also be expressed at many sites of microvascular complications including the kidney. Overexpression of angiotensinogen has been shown to cause tubular necrosis in the rodent kidney. Most of the downstream pathways of local RAAS activation have focused on superoxide as an effector molecule. Angiotensin I can also be hydrolyzed by neutral endopeptidase (NEP) to Ang1-7 and smaller peptides such as Ang1-4. Ang1-7 has been shown to have both vasodilatory and pathologic effects in certain contexts [1].

Aldosterone is a hormone produced by the adrenal glands in response to changes in serum potassium concentrations and angiotensins such as AngII. Intrarenally, aldosterone acts to increase the reabsorption of sodium ions and water in the distal tubule as well as facilitating the release of potassium ions into the urine for excretion, which ultimately results in elevations in systemic blood pressure. An increase in sodium also alters extracellular osmolarity increasing systemic blood pressure. Aldosterone elicits the majority of its effects via ligation to the mineralocorticoid receptor, which is the target of aldosterone antagonists which are both antihypertensive and often diuretic. Inhibition of the actions of aldosterone using inhibitors such as spironolactone exhibits direct renoprotective effects. In addition, beneficial effects of aldosterone antagonism on retinopathy in rodents [21] have been reported. However, the common side effect of hyperkalemia is a major limitation of widespread use of this class of agents in diabetic subjects with nephropathy.

AngII mediates most of its biologic effects via ligation to at least two specific receptor subtypes, the angiotensin type I (AT1) and angiotensin type 2 (AT2) receptors, which are each widely expressed at sites of diabetic complications. Ligation of AngII to the AT1 receptor seems to have pressor actions including vasoconstriction and activation of the sympathetic nervous system. Not surprisingly, AT1 KO mice tend to have lower blood pressure and are protected against the development of microvascular disease in the context of diabetes [22]. These findings support evidence where AT1 receptor antagonists demonstrated end-organ protection in diabetic complications. The downstream pathways affected by these agents include attenuation of fibrotic cytokine production and extracellular matrix accumulation as well as improved vascular permeability [1].

Although the AT2 receptor shares only 34% sequence homology with the AT1 receptor, these two receptors share remarkably similar affinities for AngII. In general, the AT2 receptor is thought to be fetally expressed at sites such as the developing metanephros which becomes the kidney but is not expressed to a significant extent in adults. There is some controversy as to the role of the AT2 receptor in the development of diabetic microvascular complications. It is postulated that the AT2 receptor appears to act as a functional antagonist to the AT1 receptor via protection against ROS mediated damage. However, there is an increasing body of data to show effects similar to those seen with the AT1 receptor in the context of diabetes. These effects include induction of cytokines such as VEGF and promotion of macrophage infiltration via RANTES and MCP-1.

**Obesity and dyslipidemia**

Hyperlipidemia changes the uptake of FFA by cells, both by passive diffusion and through protein-mediated pathways, most likely via proteins such as CD36 and members of the fatty acid binding protein (FABP) family. Indeed, changes in the expression of CD36 within the diabetic kidney have been previously reported. In addition, individuals with diabetes have increases in both circulating soluble CD36 (sCD36) concentrations and monocyte expression of sCD36. Studies have also
In contrast, however, increases in circulating adiponectin were inferred that serum A-FABP and E-FABP concentrations may be biomarkers for evaluating progressive nephropathy and associated cardiovascular risk in individuals with T2DM.

High-calorie diets are commonly high in saturated animal fat, which via increased gastrointestinal absorption may accelerate diabetic complications. It is also thought that obesity-induced neuropathy can be improved by dietary restriction of fat intake, relevant to diabetic individuals who are obese (Chapter 41). The most obvious effects of decreasing saturated fat intake would be on glycemic control where improved insulin sensitivity and secretion could influence end-organ function at sites of microvascular disease. Other compelling evidence of the role of overconsumption of saturated fats in progressive diabetic complications comes from studies using lipid lowering therapies such as statins and fibrates. There is, however, ongoing controversy as to whether restriction of dietary fat intake has beneficial effects on microvascular disease in diabetes.

The two major classes of lipid-lowering agents target either 3-Hydroxy-3-methylglutaryl-CoA (HMG-CoA) reductase (statins) or PPARα (fibrates including fenofibrate and gemfibrozil). Statins are thought to have broad-ranging benefits in diabetic individuals, while fibrates have shown inconsistent results with a number of pleiotropic effects such as anti-inflammatory actions, in addition to improving flow-mediated dilatation. Statins have also been shown to have anti-inflammatory effects [23], which are postulated to be due to their capacity to limit the production of isoprenoids which are key downstream mediators of the inflammatory response. Fenofibric acid derivatives (FIELD) have also been trialed with some success in diabetic retinopathy although it is unclear as to whether the effects of these agents are directly related to lipid lowering (Chapter 41). Indeed, one recent study has shown that the beneficial effects of fenofibrate on retinopathy in mouse models of diabetes are via its PPARα agonism [24]. In other experimental models, there is also some rationale for the use of fenofibrics and statins to treat microvascular complications.

Physical activity has also been shown to have effects on diabetic complications in animal models [25], via a reduction in circulating concentrations of a number of factors including advanced glycation end-products (AGEs), insulin and cytokines. As seen in nondiabetic individuals, regular moderate exercise can improve a number of risk factors relevant to the development of microvascular complications and may be an effective nonpharmacologic approach if compliance issues could be overcome [25].

**Adipokines**

Adipose tissue is highly secretory, releasing a number of factors which are modulated in response to hyperglycemia and viewed to influence diabetic complications. These adipokines include adiponectin, which is thought to reduce oxidative stress. Circulating adiponectin is also predictive of progressive kidney, retinal, and cardiovascular complications in diabetes. In contrast, however, increases in circulating adiponectin concentrations have also been shown to correlate with vascular complications in T1 diabetic individuals. In rodent models, elevating adiponectin concentrations attenuates, while prevention of the interaction of adiponectin with its receptors, worsens kidney disease in diabetes [19]. These findings are thought to be as a result of changes in AMP kinase.

Leptin is another adipokine thought to influence diabetic complications. Intravitreal leptin concentrations are increased in individuals with proliferative diabetic retinopathy, while leptin has been shown to stimulate ischemia-induced retinal neovascularization [26]. In addition, deletion of the leptin receptor in mice also results in autonomic neuropathy. Systemically, elevations in leptin concentrations are associated with renal disease and infusion of exogenous leptin leads to renal disease in various experimental models [1]. This could be partly because leptin has been reported to be pro-inflammatory, albeit in other contexts. However, it is unclear as to why leptin has specific effects at sites of diabetic complications which appear independent of its profound effects on appetite.

**Protein modifications and turnover**

One of the most complex processes that occur within cells is the folding of proteins, where stoichiometry dictates the final structure aided by posttranslational modifications such as advanced glycation, glycosylation, and phosphorylation. Indeed, changes in the functional properties of proteins are postulated as major contributors to chronic diseases including diabetic complications, which share the pathologic feature of aggregated misfolded protein deposits. This suggests the exciting possibility that “protein-misfolding” may present a common target for therapeutic intervention in diabetic complications [27].

**Autophagy**

Autophagy is a cellular process that recycles amino acids during times of need, including starvation and metabolic disorders such as diabetes [28]. These amino acids are primarily fed into the mitochondria to produce ATP via oxidative phosphorylation. Therefore, changes in autophagy could facilitate the use of inappropriate fuels for energy production. In diabetic microvascular disease, the study of autophagy is a relatively recent area of research. The growth factor, hepatocyte growth factor (HGF) has been shown to ameliorate diabetic vascular complications, which is thought, at least in part, to be via autophagic clearance of proteins and amino acids irreversibly modified by advanced glycation [1]. With respect to diabetic neuropathy, there is one study which shows that exposure of a neuronal cell line to sera from individuals with T2DM and neuropathy leads to the formation of autophagosomes and the expression of the autophagy-associated protein, beclin-1 [1]. In addition, there is evidence that the serine threonine kinase target of rapamycin (mTOR) can regulate autophagy in renal cells. Furthermore, recent studies have shown that either deletion or upregulation...
of components of the mTOR pathway, including the mTOR complexes, mTORC1, and mTOR2 [29,30] contribute to the development of diabetic nephropathy. These studies suggest that changes in autophagy may contribute to structural and functional decline at sites of diabetic complications via disruption of cellular “spring cleaning” of damaged proteins. This exciting area of research warrants further investigation.

**Posttranslational modifications of proteins**

**Enzymatic glycosylation**

Glycosylation is a critical enzymatic posttranslational modification resulting in the addition of glycans onto proteins, lipids, and other organic molecules. N-linked glycosylation, which most commonly occurs within the endoplasmic reticulum (ER), is important for the folding of a number of proteins affecting their trafficking and secretion and retention in apical membranes. ER-stress, which occurs as a result of accumulation of misfolded proteins which ultimately may overwhelm protective pathways leading to cellular apoptosis, has been shown as a common pathologic feature of diabetic microvascular complications [31]. In this context, ER stress is thought to be initiated by a number of important pathologic pathways such as advanced glycation and angiotensin II. There is some discordance, however, as to whether oxidative stress specifically is able to initiate ER stress, independently of glucose. Indeed, activation of the ER stress response occurs within the diabetic kidney or in isolated renal cells under high glucose conditions [1]. Furthermore, microarray studies using human renal biopsies from individuals with diabetes have demonstrated higher expression of specific ER stress-associated proteins including BiP (HSPA5), calnexin, and XBP-1.

Some therapies such as intravitreal injection of a purified recombinant adeno-associated virus vector (rAAV2)-58-kilodalton inhibitor of protein kinase (P58(IPK)) protect against diabetic retinopathy via improvements in endoplasmic reticulum stress. Furthermore, inhibition of ER stress ameliorates retinal inflammation in both diabetic- and oxygen-induced retinopathy mouse models. These findings suggest that ER stress is an important mediator of retinal inflammation in diabetes [1].

**Advanced glycation**

Advanced glycation of free amino groups on proteins and amino acids is a nonenzymatic posttranslational modification first discovered in food chemistry more than 100 years ago by the physician Louise Camille Maillard [32]. This reaction is influenced by many factors and is as heterogeneous as the complex biochemical reactions by which these advanced glycated products can be formed. This involves the formation of Amadori products, glycolysis, TCA cycle, the pentose phosphate pathway, glyceraldehyde-3-phosphate, and generation of the reactive carbonyl methylglyoxal. Physiologically, advanced glycation is postulated as an evolutionary pathway for labeling of senescent cellular amino acids for their recognition and ultimate turnover but it is likely that this is an over-simplistic view of these complex modifications. Indeed, recent evidence has shown that advanced glycation may modulate insulin secretion [33] and signaling [34,35], although the ultimate influence of this particular action on diabetic complications is yet to be determined. In addition, advanced glycation is thought to stabilize extracellular matrix proteins via cross-linking. It is therefore likely that these posttranslational modifications have other as yet undiscovered physiologic roles. In diabetes, not only do long-lived proteins become more heavily modified but short-lived proteins are also altered by advanced glycation.

The consequences of excessive modification of proteins by advanced glycation are numerous. Extracellular generation of AGEs alters matrix–matrix, cell–cell, or matrix–cell interactions. This has been shown under pathologic conditions to excessively cross-link the matrix resulting in stiffening and slow turnover rates. Intracellular AGE modification of extracellular matrix proteins may also alter their secretory properties and folding. In particular, modification of collagens including type IV collagen, a basement membrane glycoprotein, has been shown to alter cell adhesion thereby changing physiologic protein interactions [1].

Intracellularly, advanced glycation can directly alter trafficking and protein function. Altered production of glycolytic or Kreb's cycle intermediates or reactive oxygen species, would also lead to modification of proteins by advanced glycation [12].

AGE-modified proteins may also interact with cellular receptors. There are many AGE receptors [1], but the most widely scrutinized in diabetic microvascular complications is the receptor for advanced glycation end-products (RAGE). RAGE has many other ligands including HMGB-1, S100 calgranulins and beta-amyloid. Physiologically, the RAGE role is critical for amplification of immune and inflammatory responses, and is highly expressed on mucous membranes. The ligation of AGEs to RAGE also results in ROS generation via NAD(P)H oxidase and mitochondrial pathways. Soluble RAGE (a decoy receptor for RAGE) or RAGE neutralizing antibodies have also shown protection against complications in experimental models, findings which are also seen in RAGE-deficient mice. Conversely, transgenic overexpression of RAGE worsens kidney disease in diabetic mice [1].

Advanced glycation can activate common downstream pathways which contribute to fibrosis via excess accumulation of extracellular matrix proteins, most likely induced via RAGE. Specifically relevant to diabetic complications, AGEs can induce the production of chemokines such as monocyte chemoattractant protein (MCP-1) as well as pro-fibrotic cytokines and growth factors including transforming growth factor-β1 (TGF-β1), connective tissue growth factor (CTGF), and the angiogenic growth factor, vascular endothelial growth factor (VEGF).

Evidence to suggest a pathologic role for advanced glycation in diabetic complications primarily comes from rodent studies, which have clearly shown the efficacy of AGE-lowering
Therapies such as pyridoxamine, thiamine, alagebrium chloride, and OPB-9195 as well as lowering AGE dietary intake in averting and retarding experimental diabetic nephropathy [1]. The contribution of dietary AGES to the development and progression of complications remains controversial.

Manipulation of the enzyme glyoxalase-1, which is responsible for the removal of the AGE precursor methyglyoxal (MGO), also lowers the tissue accumulation of AGES. Indeed, lowering MGO via manipulation of glyoxalase-1 activity has functional and structural benefits for diabetic neuropathy [36] and retinopathy. Therefore, approaches that increase the activity of glyoxalase-1 or decrease the accumulation of methylglyoxal may also be worth considering as therapeutic targets in diabetic microvascular complications.

Phosphorylation
The key intracellular second messenger, protein kinase C (PKC) is a family of enzymes including at least 11 isoforms, which have been classified into three groups: (i) the conventional group which includes PKC-α, β1, β2, and γ, (ii) the novel group, and (iii) the atypical group. PKC can be activated by glucose as well as other stimuli characteristic of the diabetic milieu such as AGES and Angiotensin II. The best studied in diabetic microvascular disease is PKC-β [37]. Preclinical studies [37] have also been performed with the relatively selective PKC-β inhibitor, ruboxistaurin which inhibits both the PKC-βI and -βII isoforms. These studies defined that one of the key molecular events in the diabetic kidney that appeared to be PKC-β dependent was renal TGF-β1 expression.

A series of mice with deletions of individual isoforms of PKC have also been studied. Mice with a deletion of PKC-β after induction of diabetes, did not develop renal hypertrophy [37], or upregulation of proteins which compose the extracellular matrix including collagen and fibronectin as a result of reduced expression of the key prosclerotic growth factors TGF-β and CTGF. Oxidative stress is also attenuated in the kidneys of these diabetic mice [37]. Consistent with the major effect of PKC-β being via a TGF-β1 dependent pathway, no decrease in urinary albumin excretion was observed, a phenomenon similar to that seen in diabetic rodents treated with a TGF-β1 neutralizing antibody.

In contrast to the PKC-β KO mice, the PKC-α KO mice in response to diabetes have more prominent effects on urinary albumin excretion in addition to a decline in renal VEGF expression, as well as restoration within the glomerulus of the podocyte specific protein, nephrin [37]. This is a highly relevant finding since nephrin is strongly implicated in the pathogenesis of proteinuria including in the diabetic setting with nephrin gene deletion, nephrin gene mutations, and acquired nephrin deficiency as seen in diabetes with all these alterations in nephrin associated with increased proteinuria.

Mitogen-activated protein kinases (MAPK) including p38, initiate an intracellular cascade in response to stimuli such as cytokines, and are thought to be integral mediators of cell differentiation and apoptosis each important for the development of diabetic complications [38]. Indeed, a range of MAP kinases have been examined in the diabetic microvascular setting including p38 MAP kinase. Initially, in vitro studies demonstrated that mechanical stretch in mesangial cells leads to p38 MAP kinase activation ultimately resulting in enhanced TGF-β1 and fibronectin expression [38]. Subsequently it was shown that reactive intermediates, such as the AGE intermediate methylglyoxal, which are increased in diabetes could activate p38 MAP kinase. Furthermore, certain effects of AGES also appear to involve this signaling pathway [1]. With respect to human diabetic nephropathy, there is increasing evidence demonstrating enhanced expression of phospho ERK and p38 MAP kinase in a range of renal cell populations including mesangial cells, podocytes, endothelial cells, proximal tubular cells, and mononuclear cells within the interstitium. p38 Inhibitors have been administered to diabetic rats where effects on intrarenal hemodynamics and blood pressure were demonstrated. In addition, the p38 inhibitor, SB203580, attenuates glucose-induced tubular cell apoptosis [39]. It remains to be determined, however, if targeting p38 MAPK has benefits on the long-term renal functional and structural manifestations of diabetic nephropathy.

In the diabetic retina glucose-mediated effects on retinal pigmented epithelial cells are mediated by p38 MAPK and ERK. In addition, in models of experimental diabetes, inhibition of p38 MAPK improves retinopathy and sensory nerve function [40]. Furthermore, diabetic mice with a deficiency in MKK3, which is an upstream kinase of p38 MAPK, do not develop microvascular disease.

c-Jun N-terminal kinases (JNKs) also belong to the mitogen-activated protein kinase family and as such are responsive to stress stimuli, facilitate apoptosis and T-cell differentiation. JNK signaling has been reported as being elevated in both human and experimental diabetic complications, in particular nephropathy. In experimental models, however, inhibitors of JNK or mice with genetic deficiencies in JNK1 or JNK2 [41,42] exacerbate urinary albumin excretion and worsen the integrity of the glomerular filtration barrier in the diabetic kidney. These data contrast with other models of progressive renal disease. In a model of nondiabetic retinopathy, retinopathy of prematurity, retinal neovascularization is also improved with JNK inhibition [43]. However, the role of JNK in the development of diabetic retinopathy remains to be elucidated.

Oxidative imbalances
Nicotinamide adenine dinucleotide phosphate (NAD(P)H) oxidase
NAD(P)H oxidase [44] was originally thought to only function in neutrophils to generate vast quantities of superoxide (O₂⁻) by electron transport to destroy microbes and infected cells.
Generally, it is composed of membranous and cytosolic components and a GTPase, rac1 or rac2. Within the membrane, p22phox and Gp91phox are the major isoforms identified. Gp91phox has a number of tissue specific homologues which include Nox-5 (humans only), Nox-1, and Nox-3. The cytosolic specific subunits are p47phox, p67phox, as well as p40phox (PMID: 23379642).

Within the kidney, particular subunits are thought to be important for function. These include Nox-4, originally identified in renal tissues, which is a unique subunit in that it does not require the other subunits to generate $O_2^-$. Although originally discovered in the cytosol, Nox-4 has been recently discovered in the mitochondria.

In contrast to what was initially thought, nonphagocytic cells also express NAD(P)H oxidase [45], but the capability for production of $O_2^-$ is significantly less than in immune cells such as neutrophils. This is likely due to their differential physiologic roles, where white blood cells use NAD(P)H oxidase as a killing mechanism while in nonphagocytic cells the ROS generated are postulated to act as second messengers. Cytokines and hormones such as angiotensin II and AGEs when they ligate to their receptors rapidly activate NAD(P)H oxidase. In diabetic mice deletions of the specific NAD(P)H oxidase subunits or treatment with antisense oligonucleotides lead to improved end-organ function [44]. Although NAD(P)H oxidase may contribute to the pool of excess ROS, there are some major challenges to be overcome when targeting this pathway including redundancy among Nox isoforms and preservation of intracellular $O_2^-$ generation for use as either second messengers or alternatively in host-pathogen defense.

**Nitric oxide synthase**

Nitric oxide is produced frequently as a cellular signaling molecule by numerous cell populations. The enzyme responsible for its production from NADPH, L-arginine and oxygen often in the presence of cofactors such as bihydrobiopterin ($BH_4$) and flavin adenine dinucleotide (FAD), is nitric oxide synthase (NOS), which has multiple isoforms, inducible (iNOS), neuronal (nNOS), and endothelial (eNOS). In addition to cell signaling, NO is a powerful vasodilator and is generally thought to be vasoprotective in the context of diabetes [46].

In diabetes, there is previous evidence that uncoupling of NOS due to restriction of L-arginine availability is a major source of superoxide at sites of diabetic complications, which is produced in preference to NO in that context. There is, however, some controversy as to the contribution of NO uncoupling to diabetic complications. This is due to the temporal changes seen with NO over the course of diabetic complications. For example, early in disease development, NO production within tissues is thought to increase as a result of changes in NOS activity and therefore it has been postulated that therapeutic blockade of this pathway could be beneficial at this time. Indeed, a deficiency in iNOS or pharmacologic inhibition of NOS improves nerve conduction velocity in animal models of diabetic neuropathy [1].

Later in the disease, it is thought that functional decline in complication-prone organs is seen in conjunction with a state of progressive NO deficiency. These deficiencies in NO production could be the result of many pathways, including glucose and AGE quenching and inhibition and/or posttranslational modification of NOS. In support of this, several studies show that chronic NO inhibition has no effects or detrimental outcomes for renal disease as a consequence of diabetes [1]. This is also the case for blockade of nNOS in experimental diabetic neuropathy, where nNOS-deficient mice are not protected against diabetes-induced loss of sensory perception and intra-epidermal nerve fiber loss.

**Mitochondrial sources**

The excess generation of superoxide ($O_2^-$) by dysfunctional mitochondria has been postulated as an early event in diabetic complications [47]. This is because within mitochondria, more than 90% of oxygen in humans is metabolized during oxidative phosphorylation which utilizes glucose metabolites and other fuels that donate electrons to reduce molecular oxygen, resulting in ATP generation. Under physiologic conditions, about 1% of oxygen is only partially reduced to $O_2^-$, instead of fully to water. There are two major sites where electron leakage can occur to produce superoxide within the mitochondria, at NADH dehydrogenase (Complex I) and at the interface between Coenzyme Q (CoQ) and complex III. Therefore, based on *in vitro* studies, it has been hypothesized that excess production of $O_2^-$ is via the premature collapse of the mitochondrial membrane potential so that electrons leak to form $O_2^-$ and then $H_2O_2$ occurs rather than ATP production. However, there has never been substantiation of this phenomenon in *in vivo* models of diabetes. Despite this, there are a number of studies demonstrating mitochondrial functional abnormalities at sites of diabetic complications [12].

This is part of the therapeutic rationale for the use of antioxidants which may target the mitochondrial production of superoxide to treat diabetic complications. Idebenone is an antioxidant with preferential mitochondrial uptake by organs such as neurons, kidney, and cardiac tissues. This compound is clinically used in human respiratory chain diseases such as Friedreich Ataxia where mitochondrial generation of ATP appears to be preserved [48], particularly in cardiac tissues. Administration of exogenous Coenzyme Q has shown therapeutic benefits in animal models of diabetic complications [49,50]. A third agent, Mito Q, under investigation for the treatment of Alzheimer's disease in humans, is selectively taken into mitochondria as the result of a lipophilic triphenylphosphonium cation (www.antipodeanpharma.com). There have been studies determining the therapeutic potential of MitoQ to decrease vascular complications in experimental models of T1DM [51] and this is an area of research warranting further attention.
Antioxidants
Humans and rodents have a number of antioxidant systems that specifically control tissue ROS concentrations. Superoxide dismutase is the first of these, catalyzing the conversion of superoxide to hydrogen peroxide, consisting of three major isoforms, copper zinc superoxide dismutase (CuZnSOD, SOD1), manganese SOD (MnSOD, SOD2), and extracellular SOD (SOD3). Hydrogen peroxide is then converted to water by antioxidants such as glutathione peroxidase (GPx) and catalase. There are numerous other important cellular antioxidants, such as glutathione and numerous vitamins [12].

It is well appreciated that there are changes in the expression and activity of antioxidant enzymes at sites of diabetic microvascular disease [12]. Mice with genetic manipulations of various antioxidant enzymes such as GPx-1, however, do not develop microvascular disease despite having increased tissue hydrogen peroxide concentrations, although they do appear to be more susceptible to atherosclerosis. This is most likely due to redundancy with other GPx isoforms. Catalase overexpression is, however, protective in experimental models of type 2 diabetic nephropathy. Modulating antioxidant activity as a potential therapy for diabetic complications remains disappointing in clinical trials, however, despite some benefits of agents such as α-lipoic acid on pain and muscle weakness in diabetic polyneuropathy [1].

Inflammation
An intact acute inflammatory cascade is critical for innate immunity, and is triggered in response to a real or perceived threat to tissue homeostasis. Although the innate immune response is relatively nonspecific, adaptive immunity allows the human body to recognize and remember pathogens resulting in the capacity for an enhanced inflammatory response upon subsequent exposure to the same pathogen. Acute inflammation is a tightly regulated cascade of factors such as pro-inflammatory cytokines (e.g. tumor necrosis factor TNF-α and interleukins), chemokines (Chemokine C-C motif ligand-2 and 5; CCL2 and CCL5 and fractalkine CX3CL1), and adhesion molecules (e.g. E-selectin; intercellular adhesion molecule, ICAM-1 and vascular cell adhesion molecule, VCAM-1), that initiate the interaction between leukocytes and the endothelium to facilitate leukocyte migration towards infected or injured tissue. While acute inflammation as part of innate and adaptive immunity is beneficial, excessive or uncontrolled inflammation can promote tissue injury. Chronic inflammation is postulated as a characteristic feature seen at sites of diabetic complications.

Adhesion molecules
Hyperglycemia, hypertension, and dyslipidemia each contribute to activation of the endothelium leading to inflammation via a number of pathways including oxidative stress, NF-κB activation, dysregulation of nitric oxide synthase (NOS), and formation of advanced glycation end-products (AGEs). Not surprisingly, activation of the endothelium is a common manifestation seen in diabetes [52] where ICAM-1, VCAM-1, and E-selectins are expressed recruiting leukocytes to sites of diabetic complications. In the diabetic context, deletion of ICAM-1 protects against the development of renal disease in both experimental type 1 [53] and type 2 mouse models [54], which is also seen for retinopathy where blockade of ICAM-1 is protective against blood – retinal barrier breakdown, capillary occlusion, and endothelial cell damage [55,56].

Soluble isoforms of VCAM-1 and ICAM-1 can also be liberated by activated endothelial cells resulting in activation of leukocytes and their chemotaxis to damaged tissue sites. Increased circulating levels of sVCAM-, sICAM-1, and E-selectin are closely associated with both their increased surface expression on endothelial cells and with diabetic renal, retinal, and macrovascular complications in humans. There is also some evidence linking the levels of sVCAM-1and sICAM-1 to diabetic neuropathy [1].

Inflammatory cell infiltration
Phagocytic cells such as monocytes and macrophages are often the first infiltrating cells that arrive at sites of diabetic complications [57,58] in response to chemotactic molecules. A number of rodent studies have suggested a causal role for monocytes and macrophages in the development of diabetic complications [59]. In addition, blockade of the production of chemotactic molecules such as CCL2 is also beneficial in preventing diabetic complications in rodent models. In models of experimental diabetes, a deficiency in CCL2 impedes renal monocyte and macrophage accumulation and improves diabetic renal injury. Of course, these effects need to be considered in the context of a putative metabolic effect of CCL2 in addition to other inflammatory molecules on insulin sensitivity.

In experimental diabetes, changes in the signaling of CCL-2 which alters the actin cytoskeleton, affects nephrin expression in glomerular podocytes contributing to albuminuria. These changes in cytoskeletal rearrangement are also likely to be important for other cell types affected by diabetes. Furthermore, administration of pharmacologic antagonists of the CCL-2 receptor, CCR-2, in experimental models of diabetic nephropathy, reduced renal hypertrophy and macrophage infiltration within renal glomeruli [60,61].

Upon arrival and activation within damaged tissues, monocytes and macrophages facilitate the chemotaxis of other leukocytes such as T cells via secretion of a number of factors including interleukin -1β (IL-1β) and CCL5. Although experimental studies have shown that depletion of T-cell populations at sites of vascular injury is beneficial for atherosclerosis, other studies have shown that depletion of both β- and T-cell populations using diabetic Rag1 KO mice does not influence the development and progression of diabetic nephropathy [62]. Therefore, it seems that this area of research requires further clarification.
**Inflammatory cytokines**

Cytokines are complex interactive molecules affected by factors such as cell type, timing, and the context of their expression. These molecules are able to share receptors and act synergistically to amplify their effects and therefore it is likely that they may be difficult to rationally target therapeutically. Interleukin-1 is a cytokine released by immune cells, resident monocytes, macrophages, adipocytes, and other cells at sites of diabetic complications. The primary role of this cytokine is the recruitment and activation of other leukocytes by enhancing the expression of adhesion molecules. Indeed, there is some evidence that inhibition of IL-1β may represent a safer alternative than the more commonly targeted growth factor, vascular endothelial growth factor, in retinal degeneration [1].

The release of IL-1β also has a number of other effects on cells, including secretion of prostaglandins which alter local hemodynamics. In diabetic neuropathy, IL-1β has been interrogated more closely where it has been shown to induce nerve damage and miscommunication between Schwann cells and axons [1].

Interleukin-6 is a cytokine which can mediate cell proliferation, endothelial cell permeability and matrix overproduction. Both IL-6 and tissue necrosis factor (TNF)-α have been shown to influence glial cell and neuron behavior; these findings relevant to both diabetic neuropathy and retinopathy in humans. TNF-α signaling has also been targeted using the pharmacologic agent pentoxifylline to improve kidney function [63]. Given that TNF-α has a major role in the regulation of immune cell function, therapies targeting this axis would need to be extensively tested including detailed assessment of safety.

**Nuclear transcription factor Kappa B (NF-κB)**

NF-κB is a dimeric transcription factor commonly composed of p50 and p65 subunits [64], which is sequestered in the cytosol by the inhibitor of NF-κB, I-κBα. Following phosphorylation of I-κBα by 1κB kinase β (IKKβ), NF-κB dimers undergo nuclear translocation. The p65 subunit is thought to be a transcriptional activator controlling genes such as angiotensinogen, cytokines, and adhesion molecules in the diabetic environment [64,65]. Nuclear translocation of NF-κB is also facilitated by hyperglycemia, excess ROS, iNOS activation, and AGEs [1]. We have demonstrated that NF-κB plays a role in early renal macrophage recruitment and infiltration in the diabetic kidney [66,67]. It is thought that in diabetic vascular complications, NF-κB is a central node which may control hemodynamic and glucose-dependent signaling pathways.

The compound Bardoxolone methyl, has demonstrated potential benefits on chronic kidney disease primarily via improvements in estimated GFR (as discussed in Chapter 41). It is thought that Bardoxolone activates the Kelch-like ECH-associated protein 1 (KEAP-1)-Nrf-2 pathway, where KEAP-1 regulates NF-κB by controlling the ubiquination and breakdown of IKKβ. This is supported by research which shows that depletion of KEAP1 results in the accumulation and stabilization of IKKβ with subsequent upregulation of NF-κB-derived factors. KEAP-1-Nrf-2 pathways have also been shown to regulate ROS production in cardiac cells which may also be relevant to other sites of diabetes complications where changes in NF-κB are seen. However, recent preclinical studies suggesting nephrotoxicity [68] as well as the recent premature termination of a major clinical trial with this agent (http://www.reatapharma.com/investors-media/news/news-timeline/2012/company-statement-termination-of-beacon-trial.aspx) has reduced enthusiasm for this drug as a renoprotective agent.

Pyrrolidine dithiocarbamate (PDTC) is an experimental NF-κB inhibitor which has been used in both diabetic [66] and nondiabetic animal models where it has shown benefits on renal function, but this drug’s toxicity profile has inhibited its direct translation to the clinical setting. It is appreciated, however, that a number of therapeutics affect the nuclear translocation of NF-κB, including metformin, aspirin, vitamin B derivatives, carnosine, and thiazolidinediones [1]. However, approaches to inhibit NF-κB have not been explored fully in diabetes in humans, most likely as a result of the intimate involvement of this transcription factor in a number of essential cellular processes including apoptosis and host-pathogen defense.

**Conclusion**

As is highlighted within this chapter, the complexity of pathways involved in the pathogenesis of diabetic microvascular complications is immense. There is no doubt that diligent control of glycemia and blood pressure has had a positive impact on the morbidity and mortality associated with diabetes in most developed nations. However, given the rise in the number of cases of diabetes, in particular from developing nations, the worldwide rates of diabetic microvascular complications are set to rise dramatically. Hence it is critical to delineate the mechanisms that lead to disease development and progression and how these occur over time. Furthermore, we need to understand the complex relationship between glucose handling by sites such as the liver and pancreatic islets and how this impacts upon the pathogenesis of microvascular complications. This requires the use of rodent and other animal models where we can understand the complex relationships between pathways. The most important step, however, is to apply these findings to the clinical development of disease to discern their importance in this context given that some factors may not translate directly across to human biology. We must also not forget to look to other diseases for a greater understanding of the complex machinery that makes the human body function and how this goes wrong to produce diabetic microvascular disease.

**References**


Diabetic retinopathy and other ocular complications

Gavin S. Tan and Tien Yin Wong

1Vitreo-Retinal Service, Singapore National Eye Centre; Singapore Eye Research Institute, Duke-NUS Graduate Medical School, Singapore
2Singapore Eye Research Institute, Singapore National Eye Centre, Department of Ophthalmology, National University of Singapore & National University Hospital; Duke NUS Graduate Medical School, Singapore

Key points

• Diabetic retinopathy affects one third of persons with diabetes and is the most common cause of reversible visual loss in working adults.
• About 10% of persons with diabetes will have vision-threatening retinopathy, including diabetic macular edema.
• The presence of diabetic retinopathy is associated with an increased risk of other diabetic complications such as nephropathy and cardiovascular diseases.
• Chronic exposure to hyperglycemia and other causal factors (e.g. hypertension) is believed to initiate a cascade of biochemical and physiologic changes that ultimately lead to microvascular damage and retinal dysfunction.
• Optimal management of the modifiable risk factors blood glucose, blood pressure, and blood lipids is key to preventing development and progression of retinopathy.
• Regular dilated eye examinations are effective in detecting and treating asymptomatic vision-threatening diabetic retinopathy. Persons with T1DM should have a complete dilated retinal examination within 5 years from the onset of diabetes. Persons with T2DM should have a complete dilated retinal examination at the time of diagnosis. Subsequent examination and referral for treatment should be determined by the presence and severity of retinopathy.
• Laser photocoagulation can prevent visual loss in proliferative retinopathy and diabetic macular edema. Intraocular administration of vascular endothelial growth factor inhibitors have been established as useful new treatment modalities in diabetic macular edema. Surgical intervention by vitrectomy may occasionally be indicated in advanced proliferative disease.

Introduction

Diabetic retinopathy (DR) is a clinically well-defined, specific microvascular complication of diabetes, which is likely to develop to some degree in nearly all patients with diabetes mellitus over time. Although diabetes affects the eye in multiple ways (e.g. increases the risk of cataract and retinal vascular occlusions) [1], DR is the most common and specific ocular complication, being the leading cause of visual loss in working-aged adults in the developed world [2]. Of the 366 million persons with diabetes worldwide [3], one third will have diabetic retinopathy, of which a third is likely to have vision-threatening retinopathy, which includes severe or proliferative retinopathy and diabetic macular edema [4]. Diabetic retinopathy is reported to cause visual impairment in 10–20% of persons with diabetes [4–9]. Although DR blindness appears to have fallen in the developed world [6–8], the rapidly increasing number of persons with diabetes worldwide, especially in developing countries [10–12] where access to optimal medical care may be limited, has led to a continuing increase in the global burden of this disease [11,13–15].

Diabetic retinopathy is characterized by changes and lesions in the retinal vasculature including hemorrhages, microaneurysms, arteriolar and venular dilatation. These have a progressive nature eventually leading to areas of retinal non-perfusion, increased vasopermeability with retinal edema and exudates, and pathologic proliferative of intraocular blood vessels resulting in hemorrhage, tractional retinal detachment or neovascular glaucoma, all of which contribute to visual impairment and blindness. Apart from its effects on vision, the presence of DR also signifies a greater risk of life-threatening systemic vascular complications [16].

Recent epidemiologic, genetic, and experimental studies have furthered our understanding of the pathophysiology underlying DR. In addition, over the last 5 years, there have been several large-phase clinical trials that have provided contemporary data regarding evidence-based treatment strategies for DR, including systemic management and ocular treatment using...
antivascular endothelial growth factor (VEGF) agents. These have provided paradigm shifts in management of DR and diabetic macular edema.

**Prevalence of diabetic retinopathy**

The epidemiology of DR is well described in the literature. In many countries, DR is the most frequent cause of preventable blindness in working-aged adults [17]. In a major meta-analysis, Yau et al. described the global prevalence of DR based on 35 studies across the world with a total of 22,896 patients with diabetes and estimated the prevalence of any retinopathy to be 34.6% with 10.2% having vision-threatening retinopathy [4].

In the United States, a pooled analysis of data from different studies reported a prevalence rate of 40% for any retinopathy and 8% for vision-threatening retinopathy in T2DMs and 86% (42% for vision-threatening retinopathy) in T1DM [18,19]. Similarly high prevalence estimates have been reported in other countries (Figure 61.1) [20]. Amongst white Caucasian populations, there have been studies in the UK, Europe, and Australia [21–25]. Data from western Scotland showed the prevalence of any DR to be 26.7% [26]. Studies from Danish populations have shown an overall prevalence of retinopathy of 77% in men and 74% in women with T1DM [27]. Variations in the prevalence of retinopathy may be due in part to differences in population selection and in methods for assessing retinopathy.

In Australia, three large population-based studies assessed DR from a standardized grading of fundus photographs. The Melbourne Visual Impairment Project reported a retinopathy prevalence of 29.1% among persons aged 40 years or older with self-reported diabetes [28], The Blue Mountains Eye Study found a similar retinopathy prevalence of 32.4% among older persons aged 49 years and above with known or newly diagnosed diabetes [29], with signs of proliferative disease in 1.6% and macular edema in 5.5%. The Australian Diabetes Obesity and Lifestyle (AusDiab) study examined 11,247 adults aged 25 years or older from 42 randomly selected urban and rural communities [30]. Overall, 25% of participants with known diabetes were found to have retinopathy, including 2% with proliferative retinopathy. As in other studies, the prevalence of retinopathy was strongly related to the duration of diabetes, with a prevalence of 9.2% among those with a duration of less than 5 years, 23% for durations between 5 and 9 years, 33% for durations between 10 and 19 years, and 57% for those with a duration of 20 or more years.

In many Asian countries, the prevalence of diabetes has increased substantially over the past few decades [15,31–34], and there is concern about a potential diabetes epidemic in Asia [35]. In Singapore, for example, serial population surveys in 1975, 1985, and 1992 showed increasing prevalence rates of diabetes of 2%, 4.7%, and 8.6%, respectively, in the population between the ages of 15 and 69 years [33,34]. There are now increasing data on the epidemiology of diabetic retinopathy.

![Figure 61.1 Prevalence of diabetic retinopathy in various studies.](image-url)
Diabetic retinopathy and other ocular complications

in Asians [36–38]. In China, the Beijing Eye Study and the Handan Eye Study examined the prevalence of DR in urban and rural Chinese persons, respectively [36,38,39]. In the Beijing Eye Study, the prevalence of DR detected on the fundus photographs was 37.1%, with macular edema present in 5.2%, clinically significant macular edema in 2.6%, and vision-threatening retinopathy in 5.2%. Diabetic retinopathy was associated with rural region, longer duration of diabetes, diabetic medications use, and lower education status. The Handan eye study in a rural region in China shows that diabetic retinopathy is common, with prevalence rates of 43% for any retinopathy, 5.2% for macular edema, and 6.3% for vision-threatening retinopathy [36]. Like in other studies, the prevalence of retinopathy was strongly related to duration of disease. These estimates are higher than those reported in another study in Beijing amongst mostly urban Chinese residents [39], suggesting the need to target preventative efforts in rural areas of China. Based on these data, there are an estimated 9.2 million Chinese persons living in rural areas with DR, of whom 1.2 million have vision-threatening retinopathy [36]. There is thus a pressing need for appropriate screening and management of diabetes and its complications in rural China.

The Singapore Malay Eye Study [37] reported on DR in Asian Malays, the third largest ethnic group in Asia. The overall prevalence of any retinopathy was 35.0%, macular edema 5.7%, and vision-threatening retinopathy 9.0%. Independent risk factors for any retinopathy were longer diabetes duration, higher hemoglobin A1c, hypertension, and higher pulse pressure. Vision-threatening retinopathy additionally was associated with previous stroke, cardiovascular disease, and chronic kidney disease.

In Indians, the Aravind Eye Disease Survey in southern India reported a retinopathy prevalence of 27% in a population aged 50 years or older with self-reported diabetes [40], similar to the 22% prevalence reported from another population-based study in an urban population in Hyderabad, India [41]. The Singapore Indian Eye Study examining an urban population of ethnic Indians aged 40 and older, found a prevalence of 30.4% for any retinopathy, 7.2% for macular edema, and 7.1% for vision-threatening retinopathy. The independent risk factors for any retinopathy were younger age, longer diabetes duration, higher hemoglobin A1c, higher systolic blood pressure, lower diastolic blood pressure, previous stroke, and insulin treatment [42]. Thus, these population data suggest that diabetic retinopathy is common in Asian populations with diabetes, who share similar risk factors to White populations, suggesting that control of these risk factors may reduce both the prevalence and impact of retinopathy.

Incidence and progression of diabetic retinopathy

There are fewer data on the incidence and natural history of diabetic retinopathy. In the Wisconsin Epidemiologic Study of Diabetic Retinopathy (WESDR), the 4-year incidence of retinopathy was 40% [43,44], the 10-year incidence of retinopathy was 60–70% [45], and the 10-year incidence of macular edema was 15–25% [46]. There are few other long-term population-based incidence data using objective measures to detect retinopathy to compare with these findings [47–54]. In the United Kingdom Prospective Diabetes Study (UKPDS), a multicenter randomized clinical trial, the 6-year incidence of retinopathy was 41% [53]. The Liverpool Diabetic Eye Study reported an annual incidence of sight-threatening retinopathy of 0.3% in the first year, rising to 1.8% in the fifth year [54].

Time trends in the epidemiology of diabetic retinopathy

There have been some suggestions that over the past 30 years, better recognition and management of retinopathy risk factors and the institution of structured retinopathy screening programs have led to a decline in both the prevalence and incidence of moderate to severe microvascular diabetic complications.

Studies conducted in contemporary populations have suggested this to be the case. For example, data from both the UKPDS [53] and the Liverpool Diabetic Eye Study [54] show lower incidence rates for retinopathy, particularly sight-threatening retinopathy, than was reported previously in the WESDR and studies in the early 1980s [55]. A recent study assessed the age at which retinopathy was first diagnosed in a sample of patients with T1DM and showed that the median diabetes duration until the first occurrence of retinopathy was 16.6 years [56], which is longer than reported in previous studies. The meta-analysis review showed that estimates of retinopathy prevalence were about 10–20% lower in the seven later studies as compared to the initial WESDR prevalence [18]. The WESDR follow-up study showed that from 1980 to 2007, the estimated annual incidence of proliferative DR decreased by 77% and vision impairment decreased by 57% among persons with T1DM [6].

These data suggest that improvements in diabetes management and improved levels of metabolic and blood pressure control may have had a positive impact in reducing the prevalence and incidence of retinopathy in Western countries. Nevertheless, it is uncertain whether this declining trend will continue, with expectedly increasing number, duration, and lifespan of people with diabetes [57]. These trends however, may not apply to Asian populations. In the Handan Eye Study [36], the prevalence of retinopathy was 43.1%, higher than those reported in contemporary studies in Whites (ranging from 15.3% to 29.0%), Blacks (27.7–36.7%), and in Asian Indians (17.6%). In fact, the Handan Eye Study suggests that the prevalence of retinopathy in rural Chinese populations is more similar to rates seen in the initial WESDR.
Table 61.1 Risk factors for diabetic retinopathy

<table>
<thead>
<tr>
<th>Modifiable risk factors</th>
<th>Non-modifiable risk factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hyperglycemia</td>
<td>Age</td>
</tr>
<tr>
<td>Hypertension</td>
<td>Diabetes duration</td>
</tr>
<tr>
<td>Dyslipidemia</td>
<td>Ethnicity (Hispanic, South Asian)</td>
</tr>
<tr>
<td>Cataract surgery</td>
<td>Genetic predisposition</td>
</tr>
<tr>
<td>Obesity*</td>
<td>Puberty</td>
</tr>
<tr>
<td>Smoking*</td>
<td>Pregnancy</td>
</tr>
<tr>
<td>Alcohol consumption*</td>
<td>Nephropathy</td>
</tr>
</tbody>
</table>

*Risk factors with inconsistent evidence.

Risk factors

There are several important risk factors for diabetic retinopathy (Table 61.1).

Hyperglycemia, hypertension, and dyslipidemia

The three major modifiable risk factors, including hyperglycemia, hypertension, and dyslipidemia are discussed in greater detail in the section on systemic management of diabetic retinopathy. Briefly, reducing glycated hemoglobin levels by 1% through intensive glycemic control reduced the risk of DR by 15–40% [58], progression to vision-threatening retinopathy by 25% [59], need for laser photocoagulation by 25–30%, and risk of legal blindness by 16% [60]. Controlling hypertension with a 10 mmHg decrease in systolic blood pressure reduces the risk of DR progression by 34%, visual loss by 37%, and need for laser photocoagulation by 35% [61]. Epidemiologic studies have not found a consistent association between serum lipid levels and DR. However, studies have found that fibrates may be protective against DR progression [62,63].

Duration of diabetes

Epidemiologic studies have shown that the prevalence of diabetic retinopathy in persons with diabetes increases with the duration of disease. The WESDR study reported that after 20 years of diabetes, nearly 99% of T1DM patients and 60% of T2DM patients had developed DR [64,65]. Age of onset also plays an important role as DR rarely develops in children under 10 years of age irrespective of duration of disease.

Race

Racial/ethnic differences in the prevalence of diabetic retinopathy have been a major focus of recent research. Population-based studies suggest that the prevalence and severity of DR are higher in African Americans, Hispanics, and South Asians than in Whites, which is not fully explained by differences in the distribution of retinopathy risk factors [20,66–68]. For example, the UK Asian Diabetes Study showed that after controlling for retinopathy risk factors, South Asians were more likely to have diabetic retinopathy than Whites [20], a finding also supported by a recent large clinical trial [69]. Nevertheless, it is uncertain whether these apparent ethnic variations represent subpopulation differences associated with medical care, or variability in genetic predisposition to microvascular damage.

Genetic factors

There is evidence to support a genetic component for diabetic retinopathy [70,71]. Familial aggregation studies and clinical trials including the Diabetes Control and Complications Trial (DCCT) have demonstrated a heritable tendency for DR independent of shared risk factors [72]. Studies of other populations reported similar heritability for severe DR not fully explained by lifestyle or environmental factors [70]. A recent meta-analysis identified several genes (e.g. aldose reductase gene) associated with DR [71].

Gender

In women, pregnancy and puberty are risk factors for diabetic retinopathy in T1DM [73–75]. In the WESDR, diabetes duration after menarche, a marker of puberty onset, was associated with a 30% excess risk of retinopathy compared to diabetes duration before menarche [73]. Pregnancy is similarly associated with progression of diabetic retinopathy and risk of vision-threatening retinopathy both in T1 and T2DM [76–78]. Planned dilated retinal examination should be offered for patients with T1DM after puberty and for all pregnant women with pre-existing diabetes.

Cataract surgery

In patients with diabetes, cataract surgery has been reported to exacerbate the development and progression of diabetic retinopathy. Macular edema after cataract surgery is also a major cause of vision loss in diabetic patients, especially in those with pre-existing DR. Considerations for the management of the diabetic subjects with a cataract are discussed in the section on ocular management of retinopathy.

Relationship with other complications

Previous epidemiologic studies have also shown that diabetic retinopathy is associated with many other systemic and lifestyle factors, including nephropathy [79], obesity [80], alcohol consumption [81], smoking [82], as well as hematological markers of anemia [83], inflammation, and endothelial dysfunction [84,85]. However, some of these findings have been inconsistent, and the precise role of such factors in the pathogenesis of DR is not well defined.

Diabetic retinopathy reflects widespread end-organ microcirculatory damage and findings from epidemiologic studies suggest that DR predicts an excess risk of systemic vascular complications [16]. The presence of retinopathy, even in its mildest form, may double or triple the risk of coronary heart disease, heart failure and stroke, independent of other vascular...
risk factors [86–88]. These findings suggest the need for closer cardiovascular monitoring and follow-up for patients with DR [16].

**Pathogenesis**

Chronic exposure to hyperglycemia and other risk factors (e.g. hypertension) is believed to trigger a cascade of biochemical and physiologic changes that results in the microvascular damage and retinal dysfunction manifest in diabetic retinopathy (see Chapter 68).

**Biochemical changes**

Hyperglycemia is considered the key initiator of the cascade of retinal changes associated with diabetic retinopathy. Several biochemical mechanisms are thought to be associated with the pathogenesis of retinopathy through effects on cellular metabolism, signaling, and growth factors [17,89]. Implicated pathways include the accumulation of sorbitol and advanced glycation end-products, oxidative stress, protein kinase C activation, inflammation, and upregulation of the renin-angiotensin system and vascular endothelial growth factor (VEGF). Recognition of the potential roles for each of these processes has led to the development of new therapeutic agents, several of which have been or are currently being tested in clinical trials.

Protein kinase C is a key mediator in intracellular signal transduction and is involved in diabetic microvascular complications [90]. Hyperglycemia increases activation of retinal cellular protein kinase C, leading to increased expression of matrix proteins and vasoactive mediators, with adverse structural (pericyte apoptosis [91], basement membrane thickening) and functional (increased retinal vascular permeability and retinal blood flow) retinal vascular changes.

Hyperglycemia increases glucose flux through the polyol pathway, via which aldose reductase converts glucose into intracellular sorbitol, which may induce osmotic damage to retinal endothelial cells and pericytes. Aldose reductase gene had the largest number of polymorphisms associated with DR [92].

Oxidative stress is also a key process where hyperglycemia increases production of reactive-oxygen species, leading to activation of protein kinase C, formation of advanced glycation end-products, activation of the polyol pathway and VEGF production.

Prolonged exposure to hyperglycemia induces nonenzymatic glycation of proteins to form advanced glycation end-products (AGEs), which may contribute to retinal pericyte loss, microaneurysm formation, and vascular endothelial damage. AGEs mediate their effects on the microvasculature through interaction with the receptors for AGEs (RAGE). Modulation of RAGE may be a possible future means of therapeutic intervention for DR [93].

The intraocular renin-angiotensin system has been found to be upregulated in diabetes, and angiotensin II may stimulate expression of VEGF and other growth factors in retinal vascular endothelial cells and promote cell growth, proliferation, and the deposition of extracellular matrix proteins.

VEGF is a key mediator in the retinal vasculature changes involved in the development and progression of DR [94]. Increased levels of VEGF have been shown in the ocular fluids of diabetic patients [95]. In response to hypoxia, retinal endothelial cells, pericytes and pigment epithelial cells express VEGF, stimulating angiogenesis which results in neovascularization and increasing capillary permeability resulting in macula edema.

Recent work has identified new VEGF-independent pathways for DR [96,97]. Among these, erythropoietin has been shown to be a potent ischemia-induced angiogenic factor in proliferative retinopathy which acts independently of VEGF [98,99]. Inhibition of erythropoietin is remarkably effective in suppressing retinal neovascularization in animal models [98,100]. In the retina, erythropoietin is also expressed in response to stimuli other than ischemia [101], and it has been suggested that it may serve to protect the neural retina at the early stage of DR [102,103]. Therapeutic inhibition of erythropoietin as a treatment approach for DR must be balanced by its potential adverse effects on photoreceptor survival [96].

Proteomic analyses have identified other VEGF-independent pathways in DR. The vitreous level of extracellular carbonic anhydrase was markedly elevated in eyes with DR [104]. Inhibiting carbonic anhydrase activity has been demonstrated to reduce retinal vascular permeability in animal models [104]. However, it remains undetermined whether topical carbonic anhydrase inhibitors, which are commonly used to lower intraocular pressure in patients with glaucoma, could reduce the risk of DR [105].

Inflammation has been shown to play a critical role in the pathogenesis of DR [106–108]. An array of inflammatory mediators are upregulated in diabetes in response to hyperglycemia and other stresses (e.g. dyslipidemia) which trigger pro-inflammatory responses that may cause abnormal leukocyte–endothelial interactions and retinal microvascular damage. There is little evidence for a strong association between markers of systemic inflammation and risk of DR suggesting that this it is likely to be a local phenomenon [85,109].

Finally, there are new insights into retinal physiology that suggest the concept that DR cannot be attributed as purely a manifestation of microvascular damage. Neuroretinal compromise may occur early in the course of DR, and can precede the onset of microvascular changes [106]. It has been proposed that diabetes may reduce insulin receptor signaling in the retina, leading to neurodegeneration [106]. Experimental studies suggest that diabetes adversely affects the entire neurosensory retina, with accelerated neuronal apoptosis, and activation or altered metabolism of neuroretinal supporting cells [106]. These findings suggest that DR could be similar to peripheral diabetic neuropathy, but involving the retinal parenchyma. While the
complex relationship between the neural and vascular elements of retinopathy pathogenesis remains to be elucidated [106], understanding how diabetes affects the neural retina may ultimately lead to the development of neuroprotective agents as new potential disease modifiers [110].

**Retinal vascular changes**

Diabetes and diabetic retinopathy is associated with structural and functional changes in the retinal vasculature [111]. Quantitative assessment of the retinal vasculature with computer-based retinal image analysis has been used to study these changes in greater detail.

Increased retinal arteriolar caliber is found to be associated with the development of retinopathy in both T1 and T2DM [112–114]. Dilation of retinal arterioles may be an early indicator of microvascular dysfunction [115], implying impaired arteriolar autoregulation [114,115]. Based on the laws of Starling and Laplace, retinal arteriolar dilatation has been postulated to increase retinal capillary pressure, leading to capillary wall dilatation (microaneurysms), leakage (edema and hard exudates), and rupture (hemorrhages) [114]. Increased retinal venular caliber is independently associated with the progression of DR [116–120], and predicts the risk of proliferative retinopathy [116]. Proposed mechanisms underlying this association are likely to be multifactorial (e.g. retinal hypoxia, inflammation, endothelial dysfunction) [121–124]. Collectively, these findings suggest that retinal arteriolar dilatation may be an early subclinical marker of microvascular dysfunction preceding the development of nonproliferative DR, whereas retinal venular dilatation may be a marker of progression to more severe or proliferative retinopathy.

Fractal analysis has been used to evaluate the overall geometry of the retinal vascular network in diabetes [125]. A recent study showed that retinal fractal dimension, a measure of the density of the vascular branching pattern, is associated with early retinopathy in T1DM [126]. Furthermore, research investigating new dynamic retinal vascular changes has demonstrated that eyes with DR have reduced retinal vasodilation after flicker-light stimulation, a measure of endothelial dysfunction [127,128]. These imaging techniques may offer new means of assessing DR risk.

**Clinical features and assessment**

Diabetic retinopathy is defined clinically as the presence of typical retinal microvascular changes in an individual with diabetes mellitus. As the disease progresses, maculopathy (macular edema and ischemia) and neovascularization of the retina (vitreous hemorrhage, retinal detachment) and iris (neovascular glaucoma) leads to vision loss. The aim of clinical assessment is to detect these serious ocular manifestations, and in their absence, to assess the risk of progression to vision-threatening disease.

Direct ophthalmoscopy enables adequate assessment of DR signs, but this is enhanced by slit-lamp biomicroscopy with a condensing lens and indirect ophthalmoscopy. Severe DR may be present without symptomatic visual impairment, therefore visual acuity assessment while important, can be misleading. Examination of the peripheral fundus is critical, particularly in patients with T1DM, to avoid missing the presence of peripheral retinal ischemia and neovascularization. Referral of patients with sight-threatening retinopathy to a specialist in ophthalmology for properly and timely treatment is key to reducing the risk of visual impairment. For patients with newly diagnosed DR, a comprehensive systemic examination by physicians is also advisable [16].

The classic retinal microvascular signs of nonproliferative DR include microaneurysms, hemorrhages, hard exudates (lipid deposits), cotton-wool spots (accumulations of axoplasmic debris within adjacent bundles of ganglion cell axons) [129], venous dilation and beading, and intraretinal microvascular abnormalities (i.e., dilated pre-existing capillaries) (Figure 61.2(a,b)). Standard clinical classifications of DR are shown in Table 61.2 [130]. This classification aims to identify stages of retinopathy that may present a significant threat to visual acuity or confer a high risk for progression to those stages, which in turn guides follow-up duration and necessity for intervention.

The appearance of retinal neovascularization is a critical change in the progression of diabetic retinopathy (Figure 61.2(c,d)). The clinical features of proliferative retinopathy may vary from a few fine new blood vessels appearing on the retina or optic disc to an increase in the size, number, and extent of these new vessels. This may be asymptomatic in the early stages but predicts that the eye is at significant risk of severe visual loss. Eyes have high risk of proliferative DR when there are new vessels of about one fourth disc area on or within one disc diameter; or new vessels at the optic disc associated with vitreous hemorrhage; or new vessels away from the disc which are greater than half disc area in size and associated with vitreous hemorrhage. These eyes have a 1-year cumulative rate of severe visual loss, defined as visual acuity less than 5/200 of 10% and a 3- and 5-year risk of 35% and 50%, respectively. Advanced proliferative disease is marked by fibrovascular proliferation, and visual loss may occur suddenly due to vitreous hemorrhage from new vessels, or tractional retinal detachment from progressive fibrosis or may be subacute due to macular ischemia and neovascular glaucoma.

Diabetic macular edema is an important sign that is assessed separately from the stages of retinopathy (Figure 61.2(f)), as it can run an independent course. Macular edema is a major cause of visual impairment in persons with diabetes and results from the abnormal collection of extracellular fluid in the retina, observed clinically as the presence of hard exudates and retinal thickening in the macula. Since not all maculopathy results in visual impairment, the ETDRS study proposed a new term “clinically significant macular edema” (CSME), which is defined as retinal thickening that involves or threatens the center of the macular, the specific definition of which is elaborated in Table 61.2.
Investigations

Advances in ophthalmic imaging have resulted in these modalities playing important roles in the screening, diagnosis, and monitoring of diabetic retinopathy.

Retinal photography serves as a useful screening tool for DR, especially where access to ophthalmologists is limited. Studies have shown that retinal photography interpreted by trained readers has high sensitivity (61–90%) and specificity (85–97%) [131], and may guide appropriate ophthalmic referral [132]. There is recent evidence to suggest that automated computer grading systems may be effective in DR screening thereby reducing the workload of manual grading [133].

Fluorescein angiography is an established modality used in clinical evaluation of DR. Microaneurysms and increased capillary permeability are the earliest detectable changes. Focal areas of capillary nonperfusion represent retinal ischemia while enlargement of the foveal avascular zone identifies macular edema.
<table>
<thead>
<tr>
<th>Classification of diabetic retinopathy</th>
<th>Defining features</th>
<th>ETDRS scale [18,224]</th>
<th>Defining features (based on 7 × 30° field stereo photographs)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No retinopathy</strong></td>
<td>No retinal abnormalities</td>
<td>No retinopathy</td>
<td>No retinal abnormalities</td>
</tr>
<tr>
<td><strong>Mild nonproliferative diabetic retinopathy (NPDR)</strong></td>
<td>Microaneurysms only</td>
<td>Mild nonproliferative diabetic retinopathy (NPDR)</td>
<td>Microaneurysms only, or Venous loops in 1 field; retinal hemorrhages present; hard exudates or soft exudates in 1 field</td>
</tr>
<tr>
<td><strong>Moderate NPDR</strong></td>
<td>More than just microaneurysms but less than severe NPDR</td>
<td>Moderate NPDR</td>
<td>Two out of three moderate characteristics and or one of the following: microaneurysms and hemorrhages severe in 2 or 3 fields, IRMA present in 4–5 fields or venous beading in 1 field</td>
</tr>
<tr>
<td><strong>Severe NPDR</strong></td>
<td>Any of the following:</td>
<td>Severe NPDR</td>
<td>Microaneurysms and hemorrhages severe in 4–5 fields, or Venous beading definite in 2 fields, or IRMA moderate in 1 field.</td>
</tr>
<tr>
<td>1 More than 20 intraretinal hemorrhages in each of four quadrants</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Definite venous beading in two or more quadrants</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 Prominent IRMA in one or more quadrants and no signs of proliferative retinopathy.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Proliferative diabetic retinopathy (PDR)</strong></td>
<td>Neovascularization of optic disc (NVD) or elsewhere (NVE), preretinal hemorrhage, or vitreous hemorrhage.</td>
<td>Mild PDR</td>
<td>Fibrous proliferation on the optic disc or elsewhere or visible NVEs. NVE ≥1/2 disc area, or visible NVD, or Vitreous or preretinal hemorrhage and NVE &lt; 1/2 disc area. NVD ≥1/4–1/3 disc area, or NVD &lt; 1/4 disc area and vitreous or preretinal hemorrhage, or NVE ≥1/2 disc area and vitreous or preretinal hemorrhage, or Vitreous or preretinal hemorrhage, obscuring ≥1 disc area</td>
</tr>
<tr>
<td><strong>Diabetic macular edema (DME)</strong></td>
<td>Any apparent retinal thickening or hard exudates in posterior pole.</td>
<td>Macular edema</td>
<td>Any retinal thickening or hard exudates in posterior pole.</td>
</tr>
<tr>
<td><strong>Mild DME</strong></td>
<td>Some retinal thickening or hard exudates in the posterior pole but distant from the center of the macula.</td>
<td>Clinical significant macular edema</td>
<td>Retinal thickening at or within 500 μm of the center of the macula, or Hard exudate at or within 500 μm of the center of the macula with associated thickening of the adjacent retina, or A zone or zones of retinal thickening one disc diameter or larger, any part of which is within one disc diameter of the center of the macula.</td>
</tr>
<tr>
<td><strong>Moderate DME</strong></td>
<td>Retinal thickening or hard exudates approaching the center of the macula but not involving the center.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Severe DME</strong></td>
<td>Retinal thickening or hard exudates involving the center of the macula.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
ischemia. Retinal neovascularization presents as dye leakage into the vitreous. Diabetic macular edema has two main angiographic patterns: focal (from leaking microaneurysms), and diffuse (generalized breakdown of blood–retinal barrier).

Optical coherence tomography (OCT) has become an increasingly important imaging modality. It generates cross-sectional images of ocular structures by measuring the echo-time delay and intensity of reflected light, offering high-resolution, three-dimensional or cross-sectional images that closely approximate the histologic appearance of the retina [134]. OCT allows accurate and reproducible measurements of the retinal thickness, critical for monitoring progression and treatment response for diabetic macular edema. It is a more accurate and objective method of diagnosing macular edema than clinical examination, and has become an established clinical trial endpoint for treatment in diabetic macular edema. It is also useful to detect structural changes (e.g. vitreomacular traction or epiretinal membranes) that may indicate a need for surgical intervention. There is also a suggestion that it may be useful in screening for diabetic macular edema compared with nonstereoscopic retinal photography [135].

**Recommended timing and frequency of screening for diabetic retinopathy**

Regular dilated eye examinations can detect asymptomatic vision-threatening diabetic retinopathy effectively. In the WESDR, 14% of people with T1DM and 33% of those with T2DM developed DR within 5 years of the diagnosis of diabetes. Almost all the retinopathy cases in T1DM were mild, compared with the older participants with T2DM where 2% had proliferative retinopathy and 3% had clinically significant macular edema [46,136,137]. This suggests that DR screening should be performed at diagnosis and either yearly or second yearly thereafter in persons with T2DM. In persons with T1DM, baseline examinations could be extended to 5 years after diagnosis. The Liverpool Diabetic Eye Study, reporting retinopathy incidence in a large cohort of people with T2DM, suggests that a 3-year screening interval could be safe for patients without any evidence of retinopathy, though in patients with any sign of retinopathy, yearly or more frequent examination is recommended [138]. The most reliable and cost-effective means of screening for DR is with mydriatic digital retinal photography [139] and this has been implemented successfully in many countries resulting in cost-effective reduction of visual impairment [131,140]. Telemedicine-based digital retinal imaging may provide more cost-effective and efficacious screening for DR in diabetic patients.

Guidelines for screening and management at differing severity of DR are elaborated in Table 61.3. In practice, the timing and frequency of eye examinations in people with diabetes are often individualized. Examination at least annually is recommended for high-risk patients (e.g. those with longer duration of diabetes, poor systemic risk factor control), even in the absence of retinopathy [58]. In patients with pre-pubertal diabetes, it may be appropriate to begin retinopathy screening at puberty [73–75]. In pregnant women with nongestational diabetes, a comprehensive eye examination may be warranted during the first trimester, with follow-up throughout the pregnancy in presence of retinopathy [141]. Regular eye examinations also exert positive psychosocial effects on the care of patients with diabetes (e.g. education about risk factors, compliance) [17].

<table>
<thead>
<tr>
<th>Retinopathy severity [130]</th>
<th>Clinical implications</th>
<th>Management</th>
<th>Frequency of examination</th>
</tr>
</thead>
<tbody>
<tr>
<td>No retinopathy</td>
<td>–</td>
<td>–</td>
<td>1–2 yearly</td>
</tr>
<tr>
<td>Mild nonproliferative diabetic retinopathy (NPDR)</td>
<td>5% (within 1 year) and 14% (within 3 years) progress to PDR</td>
<td>Optimize medical therapy of glucose, blood pressure and lipids</td>
<td>Annually</td>
</tr>
<tr>
<td>Moderate NPDR</td>
<td>5–26% (within 1 year) and 30–48% (within 3 years) progress to PDR; 1.2–8.1% progress to high-risk PDR [225]</td>
<td>Refer to ophthalmologist. Optimize medical therapy of glucose, blood pressure and lipids</td>
<td>3–6 monthly</td>
</tr>
<tr>
<td>Severe NPDR</td>
<td>52% (within 1 year) and 71% (within 3 years) progress to PDR; 14.6–45.0% progress to high-risk PDR</td>
<td>Consider panretinal photocoagulation for patients with type 2 diabetes</td>
<td>3–4 monthly</td>
</tr>
<tr>
<td>Proliferative diabetic retinopathy (PDR)</td>
<td>Risk of severe visual loss</td>
<td>Indication for panretinal photocoagulation – Urgent if high-risk PDR present</td>
<td>Variable</td>
</tr>
<tr>
<td>Diabetic macular edema (DME)</td>
<td>May occur at any stage of DR</td>
<td>Indication for macular laser. Consider intravitreal anti-VEGF therapy if center involving macula edema with vision loss due to DME.</td>
<td>Variable</td>
</tr>
<tr>
<td>Clinically significant macular edema</td>
<td>Need intervention to prevent visual loss</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>
Systemic treatment of diabetic retinopathy

The most effective treatment for diabetic retinopathy is prevention and this involves the optimization of systemic risk factors. Several important studies in recent years have made a significant contribution in addressing the concerns regarding systemic optimization including the target glycemic and blood pressure levels for effective prevention of retinopathy development and progression, the efficacy of various hypoglycemic and blood pressure-lowering agents, and the role of lipid-lowering agents.

1 Glycemic control

Hyperglycemia is the key initiator in the pathogenesis and development of diabetic retinopathy. The DCCT and the United Kingdom Prospective Diabetes Study (UKPDS) are the key studies providing definitive evidence that tighter control of glycemia (HbA1c 7%) reduces the risk of development and progression of DR in both T1 and T2DM [60,142,143]. There is evidence of a small risk of initial worsening of retinopathy at the onset of therapy, but the long-term benefits outweigh this risk [144]. Each percent reduction in HbA1c (e.g. from 9% to 8%) lowers the risk of retinopathy by 30–40% and the effect is long-lasting even if subsequent metabolic control deteriorates later in life, an effect termed “metabolic memory” [145]. In keeping with these findings, a recent meta-analysis shows a graded relationship between the level of glycemia and frequency of retinopathy signs [146].

There is recent evidence to show that there is a limit to the benefit of further lowering of glycemic levels in persons with diabetes. The Action in Diabetes and Vascular Disease (ADVANCE) trial [147], demonstrated that aggressive glycemic control (HbA1c <6.5%) did not significantly impact on retinopathy development or progression in T2DM [147]. The Action to Control Cardiovascular Risk in Diabetes (ACCORD) trial showed that while intensive therapy for glycemic control did reduce macrovascular and microvascular events and progression of DR, it was also associated with increased episodes of severe hypoglycemia and increased mortality [148], although the cause of unexpected excess deaths remains unclear [149]. This suggests that in older patients and in those with established cardiovascular disease, perhaps glycemic targets should be less strict.

The Veterans Affairs Diabetes Trial (VADT) also recently reported that after 5-years follow-up, intensive glycemic control (HbA1c 6.9%; comparable to the DCCT and UKPDS) did not yield any significant benefit on the retinopathy outcomes [150]. While these findings contrast with those from the UKPDS, they could be related to the population in the VADT having more males (97%), shorter length of follow-up, and later initiation of therapy. Although it was not statistically significant, the rate of retinopathy progression in the VADT was modestly lower in the intervention group than in the control group (17% vs. 22%; p = 0.07), so that a delayed benefit of intensive glycemic control, as found in previous studies [151,152], cannot be excluded.

Recent analysis from the UKPDS and DCCT trials have suggested that the protective effect of intensified blood glucose control, especially early in disease, has a sustained effect over time. This “metabolic memory” effect persists even if glycemic control is less intensive later in the course of disease. In the DCCT, follow-up analysis showed that 4 years after the end of the trial, 17.8% of patients who were originally randomized to intensified glycemic control had progression in DR compared with 48.9% of those randomized to the conventional group; and at 10 years progression was found in 35.8% compared with 60.6% [145]. In the UKPDS, at 10 years follow-up, although the between group differences in glycated hemoglobin levels were lost after the first years, in the intensively treated group, the relative risk reduction persisted at 24% for microvascular disease compared with the less intensive group [153].

2 Blood pressure control

Epidemiologic studies and clinical trials found that hypertension is a major modifiable risk factor for diabetic retinopathy [58]. Hypertension leads to increased retinal blood flow and mechanical damage and stretching of vascular endothelial cells, stimulating the release of VEGF which contributes to progression of DR [154,155]. Each 10 mmHg increase in systolic blood pressure is associated with an approximately 10% excess risk of early DR and a 15% excess risk of proliferative retinopathy [156,157]. In the UKPDS, tighter blood pressure control reduced the risks of retinopathy progression by about one third, visual loss by half, and need for laser treatment by one third in people with T2DM [61,155]. However, these benefits are not sustainable without ongoing and long-term maintenance of blood pressure control [158].

Clinical trials have shown that renin-angiotensin system inhibitors may reduce the incidence and progression of DR beyond their blood pressure-lowering effects compared with other antihypertensive drugs. The EURODIAB Controlled Trial of Lisinopril in Insulin-Dependent Diabetes Mellitus (EUCLID) showed that lisinopril reduced the risk of retinopathy progression by 50% and proliferative retinopathy by 80% [159]. The EUCLID, however, was limited by treatment arm differences in baseline glycemia and by its short (2-year) follow-up. The Diabetic Retinopathy Candesartan Trials (DIRECT) and the Renin-Angiotensin System Study (RASS) supported these findings. In the DIRECT trials, candesartan reduced the risk of retinopathy development by 18% to 35% in T1DM, and increased regression of retinopathy by 34% in T2DM [160–162]. In the RASS, enalapril and losartan reduced the risk of retinopathy progression by 65% and 70%, respectively, in T1DM, independent of the changes in blood pressure over the period of the trial [163].

In the 10-year post trial monitoring of patients from the UKPDS, difference in blood pressure between the tight and less tight blood pressure control regimens disappears within 2 years after termination of the trial. In contrast to the sustained benefit of early glucose control, the risk reductions found during the trial for microvascular disease were no longer significant during the follow-up. This suggests that good blood pressure control must be continued if the benefit of reduced incidence and progression of retinopathy is to be maintained [158].
Lipid-lowering therapy Observational studies support a role for dyslipidemia in the pathogenesis of diabetic retinopathy [58]. The DCCT showed that the severity of retinopathy was associated with increasing triglycerides and lower levels of HDL-cholesterol [164]. A high total to HDL-cholesterol ratio and elevated LDL-cholesterol were also associated with the development of clinically significant macular edema [165]. The Fenofibrate Intervention and Event Lowering in Diabetes (FIELD) trial showed that fenofibrate, a lipid-modifying agent, reduced the need for laser treatment of vision-threatening DR by 31% in patients with T2DM over 5 years [62]. The ACCORD study reported a 40% reduction in the odds of having progression of retinopathy over 4 years afforded by fenofibrate combined with simvastatin compared with simvastatin alone, further supporting the efficacy of fenofibrate [63,166]. Interestingly, this finding did not appear to be secondary to changes in the traditional serum lipid profile [167]. Experimental studies have suggested a number of possible mechanisms for the action of fenofibrate which involve lipid and nonlipid pathways, including upregulation of apo A-I, antiapoptotic activity, oxidative stress, inflammation, protective effects on blood–retinal barrier breakdown, and neuroprotective effect [168,169]. These results suggest that fenofibrate may be beneficial in the early stages of DR.

Multifactorial intervention The Steno-2 study investigated the effect of intensive multifactorial intervention, with tight glucose regulation and the use of renin-angiotensin system blockers, aspirin, and lipid-lowering agents, in patients with T2DM and microalbuminuria [170]. The Steno-2 study encompassed treatment goals similar to those recommended in the American Diabetes Association guidelines. It showed that after 8 years of intensified, target-driven intervention that aimed to control multiple vascular risk factors the risk of retinopathy was reduced by 58% [170]. The differences in the levels of risk factors between the controlled and intervention groups were no longer significant 5 years later; however, the beneficial effects on retinopathy were sustained [170]. This is in keeping with the “metabolic memory” findings seen in the DCCT [145], and UKPDS [153] reinforcing the importance of early and strict implementation of multifactorial interventions to prevent the development and progression of DR.

Emerging medical treatments Understanding the biochemical pathways in the pathogenesis of diabetic retinopathy, various novel systemic therapies are being developed. Protein kinase C is a key mediator resulting in increased retinal neovascularization and vascular permeability. Ruboxistaurin, a well-tolerated selective protein kinase C inhibitor, has been shown in initial trials to reduce the risk of progression and need for laser treatment for diabetic macular edema [171–174], and could reduce visual impairment from long-standing macular edema [175]. However, additional trials are needed to verify these findings. Diabetes results in accumulation of AGE in the retina and the DCCT showed that AGE levels from skin biopsy may predict retinopathy progression [176,177]. A clinical trial in T1DM suggests that patients treated with Pimagedine, an aminoguanidine that inhibits the formation of AGE, has been shown in a clinical trial involving T1DM to reduce the likelihood of retinopathy progression [178].

Ocular therapy Laser photocoagulation has long been established as the gold standard for ophthalmic therapy for vision-threatening diabetic retinopathy. Timely and appropriate laser treatment is effective in preventing visual loss but has significant ocular side effects since photocoagulation is ultimately a destructive procedure. Furthermore, reversal of visual loss is uncommon even with adequate treatment. Recently, new and more effective therapeutic strategies aimed at improving vision have been developed (see Table 61.4).

Laser photocoagulation Laser therapies for diabetic retinopathy are panretinal photocoagulation (PRP) for proliferative retinopathy and macular (focal/grid) laser photocoagulation for diabetic macular edema. In PRP, the laser burns are placed over the entire retina sparing the central macula in order to promote regression and arrest progression of retinal neovascularization, possibly by reducing ischemia-driven VEGF production [95]. The Diabetic Retinopathy Study (DRS) and the Early Treatment Diabetic Retinopathy Study (ETDRS), two landmark clinical trials in the management of DR, established PRP as the primary treatment for proliferative retinopathy and severe nonproliferative retinopathy.

In the DRS, PRP reduced the risk of severe visual loss (visual acuity ≤5/200) by 50% over 5 years in more than 1758 patients with proliferative disease [179]. In the ETDRS of 3711 patients with less severe DR, early PRP reduced the risk of progression to high-risk proliferative retinopathy by 50% [180]. The DRS and ETDRS findings have been further reinforced by subsequent clinical trials [58]. The adverse effects of PRP include difficulty with light–dark adaptation (25%), a small decrease in visual acuity (10%), and peripheral visual loss (5%), which may impair night vision and affect driving [58]. Other adverse effects include changes in color vision and worsening of macular edema [58]. Both DRS and ETDRS indicate that less severe stages of DR may not benefit from laser treatment, especially given the risk of adverse effects.

In the ETDRS, macular laser reduced the risk of moderate visual loss from clinically significant macular edema by 50% [181]. The Diabetic Retinopathy Clinical Research Network (DRCRnet) showed that about 30% of patients treated with macular laser gained better vision (≥10 letters) over a 2-year period [182]. As these studies included a mixture of patients with focal or diffuse macular edema, the relative efficacy of laser treatment for specific patterns of macular edema remains undetermined.
### Table 61.4 Summary of treatments for diabetic retinopathy

<table>
<thead>
<tr>
<th>Intervention</th>
<th>Recommendation</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Diabetic macular edema (DME)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Focal laser</td>
<td>Focal laser therapy is recommended in eyes with CSME and has long-term visual acuity benefits. Treatment should be considered to DME threatening the center of the macula.</td>
<td>A, I</td>
</tr>
<tr>
<td>Intravitreal anti-VEGF agents</td>
<td>Intravitreal anti-VEGF therapy is effective as primary treatment for most cases of DME and is superior in the short term to laser treatment for visual acuity again. It is recommended for center involving DME with loss of vision secondary to DME.</td>
<td>A, I</td>
</tr>
<tr>
<td>Intravitreal steroids</td>
<td>Intravitreal steroids are not effective as primary treatment for most cases of DME, but may have a role in diffuse DME unresponsive to focal laser.</td>
<td>A, II</td>
</tr>
<tr>
<td>Vitrectomy</td>
<td>Vitrectomy may have a role in selected cases of diffuse severe DME unresponsive to focal laser, especially in the presence of vitreomacular traction.</td>
<td>B, III</td>
</tr>
<tr>
<td>Medical therapies</td>
<td>There is currently insufficient evidence to recommend the routine use of PKC inhibitors and other treatments.</td>
<td>C, III</td>
</tr>
<tr>
<td><strong>Diabetic retinopathy (DR)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panretinal photocoagulation (PRP)</td>
<td>PRP is recommended in cases with proliferative DR. Treatment should be promptly institutes if high-risk features are present. PRP is recommended in severe nonproliferative DR if any difficulty or delay in follow-up is anticipated or there are associated risk factors or signs of progression, especially in patients with type 2 diabetes.</td>
<td>A, I</td>
</tr>
<tr>
<td>Vitrectomy</td>
<td>Early vitrectomy is recommended within 3 months in patients with type 1 diabetes with severe vitreous hemorrhage and significant DR. Type 2 diabetics benefit from vitrectomy for persistent vitreous hemorrhage, although the benefit of early vitrectomy is less. Vitrectomy should be considered in eyes with severe PDR not responsive to extensive PRP, associated with traction involving the macula, or both.</td>
<td>B, II</td>
</tr>
</tbody>
</table>

### Intraocular antivascular endothelial growth factor (VEGF) agents

VEGF is a potent mediator for abnormal retinal vessel growth and leakage, making it an important therapeutic target in diabetic retinopathy [94]. Intraocular VEGF levels correlate closely with hypoxia and active neovascularization, and its levels decline after successful laser photocoagulation [95]. In addition, inhibitors of VEGF activity ameliorate ischemia-induced retinal neovascularization in animal models [94,96]. These findings all support the theory that anti-VEGF agents could arrest, or even reverse, proliferative retinopathy and macular edema.

Several agents have been evaluated in clinical trials of anti-VEGF therapy. These agents are delivered by injection directly into the vitreous of the eye (intravitreal injection), thus theoretically maximizing local efficacy and minimizing systemic adverse effects. The two most widely used anti-VEGF agents are ranibizumab, a humanized, monoclonal VEGF antibody fragment and bevacizumab, a humanized monoclonal antibody to VEGF. Newer agents are being developed including aflibercept, a soluble decoy receptor fusion protein that binds VEGF. Most trials have demonstrated benefits with the use of intravitreal anti-VEGF agents, either alone or combined with laser or intravitreal corticosteroids, for both diabetic macular edema (DME) and proliferative retinopathy or as an adjunctive therapy for vitrectomy. The RISE and RIDE trial, two phase 3 randomized trials assessing ranibizumab for diabetic macular edema found that a significantly greater proportion of patients receiving either 0.3 mg or 0.5 mg ranibizumab (33.6% to 45.7%) gained 15 letters or more in best corrected visual acuity compared with sham injection (12.3 to 18.1%) at 2 years, demonstrating a rapid and sustained improvement in vision with reduced risk of further vision loss and improved macular edema from ranibizumab with low rates of ocular and nonocular harm [183]. The RESTORE trial reported a 12-month visual acuity improvement of 6.1 letters in patients treated with ranibizumab plus sham laser and 5.9 letters in ranibizumab combined with laser therapy compared with 0.8 letters in laser plus sham injection [184]. The Diabetic Retinopathy Clinical Research Network (DRCRnet) demonstrated that patients treated with ranibizumab plus prompt laser or deferred laser had significantly better visual outcomes (9 +/− 12 letter gain) at 1 year compared with those treated with sham injections plus prompt laser (3 +/− 13 letter gain), and this difference was maintained at 3-year follow-up [185,186]. The FDA has approved monthly 0.3 mg ranibizumab intravitreal injections for DME. An expert panel recommended that ranibizumab was indicated in the treatment of DME when both center involvement and significant visual loss (20/40 or worse) due to DME were present [187]. The off-label use of bevacizumab in DME is common due to its lower cost. The DRCRnet group
examined the use of bevacizumab for DME in a phase 2 clinical trial and found that subjects who received bevacizumab alone had greater visual improvement compared with laser alone. The combination of bevacizumab with laser had no short-term benefit compared with bevacizumab alone [188]. The BOLT study compared bevacizumab with laser in DME patients with visual acuity of 20/40 to 20/320 and found at 2 years that the bevacizumab arm had a median gain of 9 letters compared with a 2.5 letter gain in the laser arm [189]. A recent systematic review examining the relative clinical effectiveness of ranibizumab and bevacizumab in DME found no statistically significant difference in the improvement in best corrected visual acuity (BCVA) of more than two lines, the mean change in best corrected visual acuity or the mean change in central macular thickness between the two drugs [190]. The role of combination therapy with both anti-VEGF injections and laser remains to be defined. In the current trials, no additional benefit has been shown with adding laser to ranibizumab. The RESTORE and DRCRnet trials may suggest that combination therapy may have inferior visual acuity outcomes compared with ranibizumab injection alone. The DA VINCI study, a phase 2 clinical trial of Aflibercept in eyes with center-involved DME, found that various regimens of aflibercept at 52 weeks resulted in a 9–13 letters gain in BCVA compared with a 1.3 letter loss of BCVA in eyes treated with laser. This benefit was also seen in the group given 8 weekly treatments after 3 initial doses suggesting that less frequent treatments may be required given the longer half-life of Aflibercept [191].

**Ocular and systemic safety of intraocular anti-VEGF therapy**

Although anti-VEGF therapy has promising clinical application in the management of diabetic retinopathy, its long-term safety in patients with diabetes has not yet been established [94]. Local adverse effects of intravitreal anti-VEGF therapy include cataract formation, retinal detachment, vitreous hemorrhage, infection, and potential loss of neural retinal cells [94].

Although injected intravitreally, significant levels of anti-VEGF agents injected may pass into the systemic circulation [94,192]. Thus, systemic inhibition of angiogenesis is a potential risk, which could compromise critical vascular responses to ischemic events in patients with diabetes. Other unwanted systemic side effects may include hypertension, proteinuria, and impaired wound healing [94], which are also of relevant concern for patients with diabetes. Prolonged systemic exposure to anti-VEGF agents, due to the lengthy half-life of some agents and the need for repeated administration, may be associated with higher risks of systemic vascular complications, such as stroke and nonocular (e.g. gastric, renal) hemorrhage [192,193]. Although clinical trials on the use of intravitreal anti-VEGF therapy for treatment of age-related macular degeneration generally show low (0.6–1.2%) incidence of stroke [194], patients with DR may be at higher risk due to pre-existing diabetes-related vascular disease [16]. A systematic review of studies examining the use of anti-VEGF therapy in DME found no consistent increase in adverse events with either ranibizumab or bevacizumab or bevacizumab although these trials were underpowered to detect differences in adverse events [190]. Meta-analysis of combined data from all ranibizumab trials of patients receiving monthly injections showed that the incidence of cerebrovascular accidents was 2.2% in the ranibizumab-treated subjects versus 0.7% in sham-treated subjects (p = 0.045) [195]. It has been suggested that ranibizumab may have a lower risk adverse effects compared to bevacizumab due to its shorter systemic half-life. The comparison in the age-related macular degeneration treatment trials found that the proportion of serious systemic adverse events was higher with bevacizumab than with ranibizumab (24.1% vs. 19.0%; RR = 1.29; 95% CI 1.01–1.66) [196]. A safety review and meta-analysis of bevacizumab and ranibizumab found that the proportion of patients with serious infections and gastrointestinal disorders was higher with bevacizumab than with ranibizumab (RR = 1.3; 95% CI 1.0–1.7), although arterial thromboembolic events were equally distributed among the groups [197]. Thus, treatment of DR with these agents must involve both clinicians and patients recognizing and assessing the potential risks and benefits.

**Intraocular steroids**

Inflammation has been shown to play an important role in the pathogenesis of diabetic retinopathy [106]. Although intraocular administration of corticosteroids has been widely used to treat diabetic macular edema [182,198,199], there has been ongoing debate regarding its use due to the potential side effects. A systematic review of seven randomized clinical trials involving 632 eyes with diabetic macular edema suggests that eyes treated with intravitreal triamcinolone, a long-acting corticosteroid, had modest improvement in visual acuity [198].

The DRCRnet conducted a multicenter randomized clinical trial in the US that compared intravitreal triamcinolone with macular laser for treating diabetic macular edema [182]. Eyes treated with intravitreal triamcinolone demonstrated a better treatment response compared with eyes treated with laser at 4 months. However, this difference disappeared after 1 year and by 2 years, eyes treated with laser had significantly better vision than eyes treated with intravitreal triamcinolone. These findings remained unchanged after 3 years of follow-up [200]. Intravitreal triamcinolone was also found to be frequently associated with significant ocular adverse effects, including secondary raised intraocular pressure and cataract progression [182,200].

The DRCRnet did not address the role of intravitreal triamcinolone in eyes that do not respond well to laser (diffuse or refractory diabetic macular edema). A recent systematic review reported that while intravitreal triamcinolone improves vision in eyes with refractory diabetic macular edema in the short term (3 months), the benefits are not long-lasting [199]. Nevertheless,
intravitreal triamcinolone may have a role as an adjunctive therapy to laser [201]. A trial on sustained release steroid implants showed some efficacy in improving visual acuity in eyes with DME [198]. Long-acting dexamethasone implants have shown improvement in vision for DME in vitrectomized eyes and for patients with persistent DME. The adverse effects may be less frequent than seen with triamcinolone but there were still a significant number of eyes developing raised intraocular pressure requiring medical treatment [202]. Sustained delivery of fluorocinolone acetonide vitreous inserts has also been shown to provide sustained improvement in visual acuity in patients with persistent DME for up 3 years compared with sham injections, although there was an increased incidence of cataract in phakic patients and glaucoma requiring incisional surgery in the fluorocinolone implant groups [203].

**Surgical intervention**

Vitrectomy is the surgical treatment for blinding complications of advanced retinopathy including persistent vitreous hemorrhage, tractional retinal detachment, and persistent macular edema with vitreous traction. It can reduce the risk of retinal neovascularization and macular edema but conversely may incite harm by increasing the risk of iris neovascularization and cataract formation [204].

The Diabetic Retinopathy Vitrectomy Study (DRVS) was the largest randomized clinical trial that evaluated the indications and timing of vitrectomy for the management of advanced proliferative retinopathy [205,206]. In the DRVS, patients with T1DM with severe vitreous hemorrhage were more likely to achieve desirable visual outcome (visual acuity 20/40) if early vitrectomy was performed (within 1 to 6 months), as compared to late vitrectomy (at 1 year). However, this benefit was not observed in patients with T2DM. This could have been related to more frequent macular ischemia in patients with T2DM. Although data from the DRVS are still valuable for clinical guidance, the threshold for performing vitrectomy has lowered in recent years due to advances in vitreoretinal surgery, including smaller gauge instruments and the ability to apply retinal laser during surgery [58].

Vitrectomy has also been suggested for diabetic macular edema refractory to laser therapy, particularly if there is evidence of macular traction (e.g., vitreomacular traction, epiretinal membrane, tractional retinal detachment close to the macula) [204]. A few trials have demonstrated some benefits of vitrectomy combined with peeling of the internal limiting membrane, the innermost layer of the retina, for diffuse or refractory diabetic macular edema [58].

**Cataract surgery**

People with diabetes have a higher risk of developing cataract requiring surgery [1]. However, cataract surgery may exacerbate macular edema [207], and retinopathy progression [208,209]. Thus, laser treatment for patients with DR prior to or promptly after cataract surgery may be useful. Clinical trials also suggest that intravitreal anti-VEGF agents during cataract surgery may be an effective adjunctive therapy to optimize visual outcome after surgery in selected cases [210,211].

**Other ocular complications**

Diabetic retinopathy may be the major cause of visual loss in diabetic patients; however, diabetes is also associated with a range of ocular disease which can also cause visual impairment and morbidity.

**Cataract**

Multiple studies have shown the association between cataracts and diabetes [212,213]. The prevalence of cataract is increased in persons with diabetes and the risk of cataract increases with longer duration of diabetes and poorer metabolic control. Deposition of advanced glycation end-products in the lens has been postulated as one possible mechanism for diabetic cataract. Cataract is a major cause of visual impairment in persons with diabetes. Although the overall outcome with cataract surgery is excellent, persons with diabetes have poorer visual outcomes compared with nondiabetic subjects after cataract surgery especially in eyes with pre-existing macular edema or proliferative DR [1]. Studies have also suggested that individuals with diabetes are at greater risk of postoperative endophthalmitis, a severe potentially blinding intraocular infection [214].

**Refractive error**

Diabetic subjects can present with fluctuating refractive error. This is due to accumulation of sorbitol in the lens as a result of hyperglycemia and increased aldose reductase activity. This results in lens swelling with change in refractive error which may fluctuate with improved glycemic control [215].

**Optic disc**

Anterior ischemic optic neuropathy (AION) is an ischemic vascular condition involving the optic nerve head which presents with acute, monocular, painless visual loss with optic disc swelling. Diabetic microvascular disease is postulated to contribute to the ischemia and up to 25% of patients with AION have a history of diabetes [216].

Diabetic papillopathy presents with mild visual loss and optic nerve swelling which should be distinguished from other causes of disc swelling such as papilledema and malignant hypertension. Diabetic papillopathy is a risk factor for progression of DR and may precede the development of AION [1].

**Ocular motility disorders**

Diabetes can cause mononeuropathies involving the third, fourth, or sixth cranial nerves, and is responsible for 25–30% of
patients aged 45 years and older who develop acute extraocular muscle palsy [217]. These patients present with binocular diplopia, and in third cranial nerve palsies, the absence of papillary involvement distinguishes it from surgical causes such as intracranial aneurysms or tumors. Recovery from diabetic cranial nerve mononeuropathies usually occurs within 3 months [218], but recurrences or subsequent involvement of other cranial nerves is common. Further investigations to exclude compressive lesions are advised in the presence of other focal neurologic deficits, progressive deterioration, or palsies occurring in young patients.

**Glaucoma**

Glaucoma is a progressive optic neuropathy associated with characteristic optic disc changes and visual field defects. Elevated intraocular pressure is the major modifiable risk factor although it may not always be present. Several epidemiologic studies have reported that primary open angle glaucoma is more common among individuals with diabetes although this finding has not been found in other studies [219]. The risk of glaucoma has been reported to be 1.6 to 4.7 times higher in individuals with diabetes compared to nondiabetics [1]. This increased risk may be related to optic nerve damage secondary to impaired microvascular circulation; or concomitant cardiovascular risk factors that affect vascular perfusion to the optic nerve head. Patients with primary angle closure glaucoma are more likely to have abnormal glucose levels than either those with primary open angle glaucoma or healthy controls. This association between primary angle closure glaucoma and diabetes may be related to systemic autonomic dysfunction or increased lens thickness secondary to hyperglycemia.

Neovascular glaucoma has been consistently associated with diabetes, with proliferative diabetic retinopathy accounting for 32–43 % of cases [1]. Neovascularization of the iris occurs secondary to ischemic DR with the release of growth factors such as VEGF. This fibrovascular proliferation disrupts normal aqueous outflow resulting in raised intraocular pressure and neovascular glaucoma. Regression of neovascularization can occur following early treatment with panretinal photocoagulation [220]; however, visual prognosis is usually guarded.

**Corneal disorders**

Diabetes is known to affect the corneal epithelium leading to corneal erosion, persistent epithelial defects, or corneal ulcers. A reduction of hemidesmosomes, which attach epithelial cells to the extracellular matrix, may contribute to weakness in the adhesion of diabetic corneal epithelium to the underlying stroma [221]. Epithelial defects and recurrent erosions have been reported to occur after photocoagulation and vitrectomy. Persons with diabetes also often have reduced corneal sensation related to peripheral neuropathy and limbal vasculopathy [222]. This deficit may predispose to bacterial corneal ulcers and complications with contact lens use.

**Other retinal conditions**

Central and branch retinal vein occlusion may occur in greater frequency in diabetic individuals although epidemiologic studies have not shown a consistent relationship [1]. Retinal vein occlusion must be distinguished from DR in persons with diabetes who present with acute visual loss and unilateral retinal changes. The prevalence of diabetes in patients with central retinal artery occlusion is reported to be much higher than in the age-matched general population [223]. Retinal artery occlusion presents with sudden unilateral severe visual loss and should be referred immediately to an ophthalmologist for management, although visual prognosis is poor even with treatment.

**Future directions**

In spite of our growing understanding of the pathophysiologic and risk factors for diabetic retinopathy, much remains unanswered. Traditional risk factors only account for a small proportion of the variation in retinopathy risk in persons with diabetes. Novel markers, pathophysiologic pathways or genetic determinants need to be identified to explain the remaining variation in risk. Although new treatments have emerged in the systemic and ocular management of DR, the most efficacious clinical application of these modalities has yet to be defined. Fenofibrate has been shown to reduce the risk of retinopathy, but further studies are needed to define the best responders in order to guideline when and to whom the medication should be prescribed. Antivascular endothelial growth factors have heralded a paradigm shift in the treatment of DME; however, the lack of sustained effect and the need for repeated injections remain unsatisfactory. The development of longer acting therapies or less invasive means of drug delivery is needed to address these concerns. Developments in telemedicine and nonmydriatic screening will allow us to improve the early detection and prevention of vision-threatening diabetic retinopathy.

**References**


Diabetic retinopathy and other ocular complications


177 Monnier VM, Bautista O, Kenny D, et al.: Skin collagen glycation, glycoxidation, and crosslinking are lower in subjects with long-term intensive versus conventional therapy of type 1 diabetes.


CHAPTER 62
Diabetic nephropathy

George Jerums1,3, Elif I. Ekinci1,3, Erosha Premaratne1,3, Scott T. Baker1,3, Sianna Panagiotopoulos1,3 and Richard J. MacIsaac2,3
1Endocrine Centre & Department of Medicine, Austin Health, Melbourne, Victoria, Australia
2Department of Endocrinology & Diabetes, St Vincent’s Hospital Melbourne, Victoria, Australia
3University of Melbourne, Victoria, Australia

Key points

• Diabetic nephropathy (DN), responsible for half the burden of end-stage renal disease (ESRD) in westernized societies, affects approximately one third of patients with T1DM or T2DM. Ninety percent of diabetes-related ESRD occurs in patients with T2DM, with this proportion likely to increase due to the obesity epidemic.

• DKD is initiated by inadequate glycemic control. Intensive glycemic control prevents the onset of DN but current evidence suggests it is less effective in later stages of DN.

• DKD progression is promoted chiefly by hyperglycemia and hypertension, with control of blood pressure assuming greater importance in later stages. In overt DN, approximately 30% of the renoprotective effect of antihypertensive therapy is attributable to renin-angiotensin system (RAS) inhibition.

• Duration of diabetes, reflecting cumulative metabolic and hemodynamic stress, is a major determinant of the onset of DN. Rapid progression of DN is associated with puberty in T1DM, and with ethnicity (nonmodifiable) and socioeconomic status (modifiable) in T2DM.

• DKD is described by elevations in albumin excretion rate (AER) and/or decreases in glomerular filtration rate (GFR). Classically, DN is defined specifically as the presence of persistent, clinically detectable proteinuria, associated with an elevation in blood pressure (BP) and a decline in GFR. However, subclinical proteinuria—termed microalbuminuria—has been recognized as a definable early stage in the natural history of rising albuminuria in DKD. Microalbuminuria is defined according to an albumin excretion rate (AER) of 20–200 μg min−1 (30–300 μg d−1) or an albumin to creatinine ratio (ACR) of 2.5–25 mg mmol−1 in males and 3.5–35 mg mmol−1 in women (Table 62.1). The development of persistent microalbuminuria—that is, microalbuminuria detected in at least two out of three consecutive 24-hour urine

collections or early morning midstream samples—has been equated to incipient nephropathy, although more recently it has also been recognized as an important risk marker of cardiovascular disease and death in people with diabetes.

The combined contributions of albuminuria and GFR to the prognosis of kidney disease, including DKD, are reflected by the development of a two-dimensional composite ranking system, based on renal and cardiovascular risk [1]. According to this classification, albuminuria is ranked in three ACR categories, normo-, micro-, and macroalbuminuria and GFR ranked in five estimated GFR (eGFR) categories (G1–5) as shown in (Table 62.2). The term DKD will generally be used throughout this chapter instead of the term DN, unless reference is specifically made to studies which have involved experimental models or subjects with hypertension and proteinuria that fulfill the criteria for the original description of overt nephropathy.

### Aetiology of DKD

In T1DM, the initiator of DKD is hyperglycemia, whereas elevated blood pressure (BP) is a progression promoter as well as a consequence of DKD. Dyslipidemia has not been shown to be an initiator or pathogenetic promoter of DKD. In T2DM, DKD is also initiated by hyperglycemia but hypertension, atherosclerosis, and aging-related intrarenal vascular disease may contribute as progression promoters.

Although subtle changes in renal function and structure may develop in most subjects with diabetes, clinically overt DKD

---

**Table 62.1** Classification of albuminuria in people with diabetes

<table>
<thead>
<tr>
<th></th>
<th>AER (µg min⁻¹)</th>
<th>ACR (mg mmol⁻¹/mg g⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Females</td>
<td>Males</td>
<td></td>
</tr>
<tr>
<td>Normoalbuminuria</td>
<td>&lt;20</td>
<td>&lt;30</td>
</tr>
<tr>
<td>Macroalbuminuria</td>
<td>&gt;200</td>
<td>&gt;300</td>
</tr>
</tbody>
</table>

Source: Adapted from Levey 2011 [61]. Reproduced with permission of Nature Publishing Group.

AER, albumin excretion rate; ACR, albumin to creatinine ratio.

---

**Table 62.2** Classification of chronic kidney disease (CKD) incorporating albuminuria category (albumin to creatinine ratio) and estimated glomerular filtration rate (eGFR) stage, proposed in the Kidney Disease: Improving Global Outcomes (KDIGO) Controversies Conference Report 2012 [62].

<table>
<thead>
<tr>
<th>CKD Classification as proposed by KDIGO Controversies Conference Report 2012</th>
<th>Albuminuria stages, description and range (mg/g:mmol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFR Stages description and range (mL/min per 1.73m²)</td>
<td>A1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stage</th>
<th>Description</th>
<th>Optimal and high normal</th>
<th>High</th>
<th>Very high and nephrotic</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>High and Optimal</td>
<td>&gt;105</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>90–104</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G2</td>
<td>Mild</td>
<td>75–89</td>
<td>60–74</td>
<td></td>
</tr>
<tr>
<td>G3a</td>
<td>Mild-moderate</td>
<td>45–59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G3b</td>
<td>Moderate</td>
<td>30–44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G4</td>
<td>Severe</td>
<td>15–29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G5</td>
<td>Kidney Failure</td>
<td>&lt;15</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: The accuracy of the albumin to creatinine ratio (ACR) in stage A2 (microalbuminuria) can be improved by allowing for the greater muscle mass in men, which leads to higher urinary creatinine concentrations. Sex-specific ACR ranges in stage A2 are as follows:

**Men**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5–25</td>
<td>3.5–35</td>
</tr>
<tr>
<td>25–250</td>
<td>35–350</td>
</tr>
</tbody>
</table>

**Women**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5–25</td>
<td>3.5–35</td>
</tr>
<tr>
<td>25–250</td>
<td>35–350</td>
</tr>
</tbody>
</table>

develops in approximately one in three Caucasian subjects with either T1DM or T2DM. By contrast, the risk of DKD in T2DM is often increased in other ethnic groups, including a 5–10-fold higher risk in Pima Indians and Indigenous Australians. In Indigenous Australians, albuminuria often precedes onset of diabetes, suggesting a different pathogenetic pathway.

Recent evidence indicates that genetic, epigenetic, and dietary factors may influence the development and progression of DKD. Some families appear more susceptible to DKD than others [2], and recent studies have linked specific gene variants with different DKD phenotypes. For instance, the gene variants associated with an increase in albuminuria may differ from those associated with a decrease in GFR in patients with T2DM and advanced DKD.

In T1DM patients with DKD, impaired GFR has been associated with several variants in genes controlling matrix metalloproteinases and with endothelial nitric oxide synthase (eNOS). In T2DM, impaired GFR has been linked with variants in the engulfment and cell mobility (ELMO 1) gene and the gene for the key proinflammatory cytokine, interleukin-6 (IL-6).

Despite these advances, both candidate gene studies and genome-wide linkage scans have so far failed to identify a major gene that explains the strong familial clustering of DKD in T1DM and T2DM. By contrast, major advances have occurred in understanding the genetics of nondiabetic kidney disease in African Americans.

First, hypertension-associated end-stage renal disease (ESRD) is substantially related to polymorphisms in the nonmuscle myosin heavy chain 9 gene (MYH9) and this may explain the poor response to blood pressure control in African Americans with hypertensive nephrosclerosis. Second, a single gene variant in the apolipoprotein L1 gene (APOL1) has been shown to fully account for excess rates of nondiabetic kidney disease in African Americans. Whether gene variants in APOL1 influence the course of DKD in African American or in other ethnic groups is not yet clear.

Epigenetic factors represent another mechanism linking genes, the environment and DKD [3]. Potential epigenetic markers for DKD progression have been identified by comparing DNA methylation in patients with T2DM, with or without DKD. Epigenetic differences leading to lower levels of DNA methylation in peripheral blood leukocytes from patients with DKD were associated with a predisposition to ESRD. However, prospective studies are required to show that epigenetic factors predict DKD.

Dietary factors may impair renal function in patients with diabetes. Obesity predisposes to focal sclerosing glomerulonephritis and may therefore contribute to a decline in GFR, especially in patients with T2DM. High-protein diets may decrease GFR in subjects with preexisting renal impairment. Exogenous advanced glycation end-products (AGEs) as well as advanced lipid peroxidation end-products (ALEs) may also contribute to renal damage in patients with diabetes [4].

The kidney is an important regulator of AGE metabolism and circulating AGE levels are markedly elevated in patients with renal insufficiency with or without diabetes. This applies in particular to low molecular weight AGEs which are not bound to proteins and may damage target cells by binding to cellular AGE receptors (RAGE) and promoting vascular dysfunction [5].

The cellular mechanisms responsible for renal damage include activation of RAGE, which triggers increases in protein kinase C (PKC), nuclear factor kappa B (NFκB), transforming growth factor-β (TGF-β), and connective tissue growth factor. RAGE activation induces glomerular matrix production and increases oxidative stress through both cytosolic and mitochondrial superoxide production [6]. Studies in experimental diabetes have also shown that RAGE activation promotes epithelial-mesenchymal transdifferentiation of renal tubular cells, thereby contributing to interstitial fibrosis.

Immunohistochemical studies in subjects with diabetic or nondiabetic kidney disease have shown that AGEs accumulate in the mesangium and the glomerular capillary wall [7]. While it is not yet clear whether AGEs contribute to disease progression as opposed to being markers of disease severity, there is preliminary evidence supporting a pathogenetic role for exogenous AGEs. In a group of diabetic subjects, short-term dietary AGE restriction was associated with a significant reduction in markers of inflammation, plasma C-reactive protein, and peripheral mononuclear cell tumor necrosis factor-α (TNF-α). Furthermore, urinary excretion of the AGEs carboxymethyllysine (CML) in T1DM and methylglyoxal in T2DM has been associated with increases in albuminuria.

In summary, the kidneys excrete AGEs and are also targets for AGE-induced injury. Long-term exposure to a diet high in AGEs (as occurs with prolonged or high temperature cooking of proteins) may therefore contribute to the rate of progression of renal impairment in the elderly with or without coexisting diabetes. The postulated contributions of oxidative stress, growth factors, inflammation, AGEs, and intraglomerular pressure to the development and progression of DKD are summarized in Figure 62.1.

Recent studies in normoalbuminuric patients have shown that circulating markers of inflammation such as TNF-α receptors (sTNFR1 and sTNFR2) predict decline in renal function in both T1DM and T2DM [8,9]. However, it is not known if TNF-α receptors are directly linked to pathogenetic mechanisms in DKD.

The high intra-individual variation of albumin excretion rate (AER) in diabetes and the finding of advanced renal structural changes in renal biopsies performed in subjects with minimal increases in AER suggest a need for new approaches to the detection of DN at an early stage. One such approach is proteomics, a method that allows detection of not only intact albumin, but also over 60 albumin fragmentation products. Cross-sectional studies have identified polypeptide patterns which have differentiated patients with or without DN using capillary electrophoresis coupled with mass spectroscopy (CEMS). A recent longitudinal study has shown that early DKD can be detected.
**Mechanisms in Diabetic Nephropathy**

**High glucose & mitochondrial & cytoplasmic ROS**

- Protein Lipids
- E/C AGEs
- I/C AGEs
- I/C Methylglyoxal
- DMG
- RAGE Endothelial
- Glyoxalase

**Intraglomerular pressure**

- Angiotensin 2
  - via AT1 receptor
  - Intraglomerular pressure

- E/C AGEs
- **E/C AGEs**
- **Vascular endothelial growth factor (VEGF)**
- RAGE, PKC, VEGF
- RAGE, PKC, VEGF
- RAS inhibition
- RAGE, PKC, VEGF
- RAGE, PKC, VEGF

**Hexosamine & polyol pathways**

- PKC
- I/C ROS
- sRAGE
- I/C ROS
- I/C ROS

**Podocyte dysfunction & loss**

- Impaired tubular albumin uptake
- Megalin, Nrf2
- Proteinuria

**Mesangial matrix expansion**

- Tubulointerstitial fibrosis
- **GFR**

* High glucose mediated afferent glomerular arteriolar vasodilator ± systemic hypertension

**Efferent glomerular arteriolar vasoconstriction**

* Figure 62.1 Mechanisms in diabetic kidney disease (DKD). This diagram summarizes postulated contributions of oxidative stress, growth factors, inflammation, AGEs and intraglomerular pressure to the development and progression of DKD. AB, antibody; AGEs, advanced glycation end-products; AMP-K, AMP activated kinase; AT1, angiotensin 2 type 1 (receptor); DAG, diacylglycerol; E/C, extra cellular; EMT, epithelial mesenchymal transformation (induced by AGEs via RAGE); eNOS, endothelial nitric oxide synthase; I/C, intra cellular; IL-6, interleukin 6; MCP-1, monocyte chemotactic protein-1; mTOR-1, mechanistic target of rapamycin; NFκβ, nuclear factor Kappa β; PKC, protein kinase C (alpha or beta); RAGE, receptor for AGEs; sRAGE, soluble (circulating) RAGE; RAS, renin-angiotensin system inhibitors (RAS inhibitors do not prevent nephropathy but do delay the progression of overt nephropathy); ROS, reactive oxygen species; TGF-β, transforming growth factor beta; SMAD, TGF-β regulated transcription factor; VEGF, vascular endothelial growth factor.

by CEMS before the development of microalbuminuria [10]. It is not yet clear whether a CEMS pattern, attributed largely to collagen fragments, or whether novel biomarkers, identified by CHIP (chromatin immunoprecipitation) technology, will be the most appropriate path towards earlier diagnosis of DKD.

**Natural history of DKD**

**Type 1 diabetes**

At diagnosis of T1DM, there are transient increases in GFR, AER, and kidney size, which return to normal levels in most patients over several months [11]. However, GFR remains high in about one third of patients. This state of prolonged hyperfiltration has been associated with later progression to nephropathy in some studies but this has not been confirmed by others. The pathogenesis and significance of hyperfiltration are discussed in a separate section in this chapter.

In patients who progress to ESRD, microalbuminuria usually develops within 5–10 years of the diagnosis of diabetes. Factors that have been associated with the development of microalbuminuria include (nonmodifiable) duration of diabetes, family history of DKD, male sex, ethnicity, (modifiable) poor glycemic control, increased blood pressure (BP), presence of retinopathy, and smoking. Thus, both metabolic and hemodynamic factors contribute to the development of DKD: in early DKD, glycemic control plays the major role [12], while in later stages, BP is the most important determinant of the rate of decline of GFR [13]. With the development of overt DN (macroalbuminuria), the rate of decline in GFR rises to around 10–12 mL min$^{-1}$ per year, which intensive BP control can reduce to around 4 mL min$^{-1}$ per year in some but not all subjects [14].

The cumulative incidence of overt DN after 20 years duration of T1DM has fallen in some studies, from 25% in subjects diagnosed between 1953–1962 to 15% in later studies. However, a recent analysis of incidence of ESRD in T1DM by decade of diagnosis has demonstrated no difference in renal outcomes over the preceding three decades.

**Type 2 diabetes**

The development of DKD in T2DM shows both similarities and differences to that in T1DM. An important difference is that microalbuminuria and raised BP often precede the diagnosis of nephropathy in T2DM, or indeed T2DM itself.
Treatment of hypertension with renin-angiotensin system (RAS) inhibitors may therefore lower AER and create a state of tonic suppression of AER in the normoalbuminuric range, in which the development of microalbuminuria would only be revealed by interruption of antihypertensive therapy.

Another difference in T2DM is that ethnicity exerts a major effect on the prevalence of DKD. Compared to European subjects, there is a higher prevalence of microalbuminuria in Asians, African Americans, Maori and Pacific Islanders, and Indigenous Australians. This is also reflected in a higher prevalence of overt nephropathy and ESRD. For instance, prevalence of ESRD is 5–10-fold higher in Indigenous Australians than in white European subjects with T2DM.

As shown in the UKPDS, at each step in the progression of AER from normo- to micro- to macroalbuminuria in T2DM, the rate of stepwise increase in AER category (2–4% of patients per year) is matched by a similar increase in risk of fatal cardiovascular events. This situation of competing risks raises the possibility that aggressive use of statin therapy since the early 1990s has decreased cardiovascular deaths, and in doing so may have allowed more people with overt DN to progress to ESRD. This, coupled with progressive increases in incidence of T2DM, may have masked a decrease in the rate of progression of DKD in T2DM following the introduction of multifactorial intervention, as first shown in the Steno-2 Study [15].

A summary of the prevalence, main clinical features, and renal structural changes in DKD is shown in Table 62.3.

**Predictive value of albuminurin below the microalbuminuric range**

It is important to appreciate that the lower limit of microalbuminuria (AER 20 μg min⁻¹, 30 mg per 24 hours, or ACR 2.5 mg mmol⁻¹ in men and 3.5 mg mmol⁻¹ in women) was based on three retrospective studies in patients with T1DM in the early 1980s. In healthy young adults, the mean AER is 5 μg min⁻¹ and rarely exceeds 15 μg min⁻¹. In elderly, nondiabetic subjects small increases in AER, short of microalbuminuria, are associated with macrovascular disease. In young subjects with T1DM, small increases in AER (10–20 μg min⁻¹) predict future development of microalbuminuria, whereas in older subjects with T2DM, borderline increases in AER predict either renal or cardiovascular outcomes or both [16]. Recent data in young patients with T2DM and features of the metabolic syndrome also suggest that increases in AER predict cardiovascular as well as renal outcomes to a greater degree than in patients with T1DM of similar age.

**Hyperfiltration**

**Pathophysiology**

Hyperfiltration (HF) has been postulated as a maladaptive response to glomerular hemodynamic disturbances early in the course of DKD [17]. At a glomerular level, it is caused by...

---

**Table 62.3** Comparison of nephropathy in T1DM and T2DM

<table>
<thead>
<tr>
<th></th>
<th>T1DM</th>
<th>T2DM</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Developing DN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>20–30%</td>
<td>20–30%</td>
<td>DN develops faster in obese adolescents than in older patients with T2DM.</td>
</tr>
<tr>
<td>Non-Caucasian</td>
<td>20–30%</td>
<td>30–60%</td>
<td>Over 90% of diabetic patients with ESRD have T2DM. Ethnicity is major determinant of ESRD in T2DM, with incident ESRD due to diabetes varying from 15–30% in Europe to 60% in Mexico.</td>
</tr>
<tr>
<td>Proportion of total ESRD attributable to diabetes</td>
<td>5–10%</td>
<td>20–40%</td>
<td>Average annual GFR decline in nondiabetic subjects aged &gt;40 years is 1.0 mL min⁻¹ per 1.73m². About one in four subjects with diabetes develop normoalbuminuric renal insufficiency (GFR &lt;60 mL min⁻¹ per 1.73m²).</td>
</tr>
<tr>
<td>Decrease in GFR according to AER status (mL min⁻¹ per 1.73m² per year)</td>
<td>Normo 0–2</td>
<td>1–4</td>
<td>Insulin resistance, obesity, and metabolic syndrome contribute to DN in T2DM. Loss of nocturnal BP dip often precedes onset of microalbuminuria in T1DM.</td>
</tr>
<tr>
<td></td>
<td>Micro 1–4</td>
<td>2–5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Macro 3–12</td>
<td>3–12</td>
<td></td>
</tr>
<tr>
<td>Associations with DN</td>
<td>± follows DN onset</td>
<td>+++ usually precedes DN</td>
<td></td>
</tr>
<tr>
<td>Insulin resistance</td>
<td>follows DN onset</td>
<td>follows DN onset</td>
<td>usually precedes DN</td>
</tr>
<tr>
<td>Hypertension</td>
<td>follows DN onset</td>
<td>follows DN onset</td>
<td>variable</td>
</tr>
<tr>
<td>Dyslipidemia (↑ TG, ↓ HDL)</td>
<td>follows DN onset</td>
<td>follows DN onset</td>
<td>often precedes DN</td>
</tr>
<tr>
<td>Retinopathy</td>
<td></td>
<td>±</td>
<td></td>
</tr>
<tr>
<td>Coronary disease</td>
<td></td>
<td>±</td>
<td></td>
</tr>
<tr>
<td>Renal structural characteristics</td>
<td>+++</td>
<td>++</td>
<td>Suspect nondiabetic renal disease if rapid onset of macroalbuminuria in absence of retinopathy.</td>
</tr>
<tr>
<td>Glomerular</td>
<td></td>
<td>±</td>
<td></td>
</tr>
<tr>
<td>(diffuse and nodular glomerulosclerosis)</td>
<td></td>
<td>++</td>
<td>Intra- and extra-renal vascular changes are prominent in older patients with T2DM.</td>
</tr>
<tr>
<td>Extraglomerular</td>
<td>±</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(interstitial fibrosis)</td>
<td>±</td>
<td>++</td>
<td></td>
</tr>
<tr>
<td>(arteriolar hyaline change)</td>
<td>±</td>
<td>++</td>
<td></td>
</tr>
</tbody>
</table>

AER, albumin excretion rate; DN, diabetic nephropathy; ESRD, end-stage renal disease (eGFR <15 mL min⁻¹ per 1.73m²); GFR, glomerular filtration rate.
increases in the glomerular capillary plasma flow rate and mean glomerular capillary hydraulic pressure, which in turn is due to changes in efferent and afferent arteriolar resistance and changes in systemic arterial pressure [18]. Increased intraglomerular pressure, as a result of greater dilatation of the afferent glomerular arteriole and/or less constriction of the efferent arteriole, is a hallmark of early DKD. Autocrine modulation of vascular smooth muscle cell growth and proliferation in response to the augmented activity of the intrarenal renin-angiotensin system and the proliferative influence of fibroblast growth factor (FGF-23) have been documented in in vitro studies [19].

A major factor that has been incriminated as a cause of HF in diabetes is acute hyperglycemia, by means of activation of tubuloglomerular feedback and vasodilatation. Resolution of HF has been achieved by re-establishing a normoglycemic state [20]. Peptide and nonpeptide vasoactive agents may also be involved in the mediation of vascular changes seen in early DKD. For instance, increased nitric oxide synthesis has been associated with HF. However, it is not clear if the changes in nitric oxide production play a causal role in the pathogenesis of glomerular HF or whether they are secondary to increased renal blood flow.

HF at a renal level is also associated with a number of clinical conditions other than diabetes, such as systemic hypertension, obesity/metabolic syndrome. HF at a glomerular level occurs with congenital or acquired reduction in renal mass. Enhanced systemic production of renin-angiotensin has been recognized as a factor causing exaggeration of the pathologic effects of the intra-renal changes described earlier. Pharmacologic agents blocking the action of the renin-angiotensin system are well proven to reduce glomerular HF [21]. High dietary sodium intake and increased proximal tubular sodium resorption, as part of a tubuloglomerular feedback mechanism has also been proposed as a mediator of HF.

**Clinical studies**

Glomerular HF is a well-characterized phenomenon seen in approximately 24–30% of patients with early T1DM [22]. With time, glomerular filtration falls gradually in parallel with increasing albuminuria, which may eventually lead to progressive renal failure (Figure 62.2). There is no universally accepted definition for HF, but it is generally defined as a GFR of >120–135 mL min$^{-1}$ per 1.73 m$^2$, dependent upon the GFR methodology and the study population. While over the past four decades a number of cross-sectional and longitudinal studies have examined the prevalence of HF in diabetes, the causative or predictive role of HF in the pathogenesis of microalbuminuria or renal impairment remains largely uncertain.

A recent meta-analysis of 780 subjects with T1DM concluded that glomerular HF increased the risk of albuminuria and progression of DKD. The relative risk of progressive albuminuria was 2.71 in those with HF versus normal GFR at baseline. HF was associated with higher HbA1c in this study. By contrast, a previous study of children with T1DM, which also demonstrated a correlation of HF with microalbuminuria, failed to demonstrate an association with metabolic control. However, the concept of HF being a predictor of microalbuminuria in T1DM is not a universal finding. A case-controlled prospective study of 10 years of follow-up of 50 subjects [23] and a more recent single-center study of 15-year follow-up of 426 subjects with T1DM [24] are examples of studies that failed to demonstrate an association between glomerular HF and subsequent renal disease defined in terms of development of micro- or macroalbuminuria.

The pathogenetic role of glomerular HF is even less clear in T2DM, due to a paucity of long-term studies. The prevalence of HF is high among patients with T2DM in certain ethnic groups, such as the Pima Indians. Most studies of HF in T2DM subjects conclude that it is mainly a phenomenon in younger subjects. However, this may be related to the fact that HF is “masked” in older subjects due to the age-related decline in renal function. The prevalence of HF in T2DM has been found to increase from 7.4% to 16.6% after adjusting for the expected age-related decline in GFR [25].

In longitudinal studies, HF in T2DM has been associated with a greater rate of decline in GFR compared to T2DM subjects with normal GFR and non-diabetic controls over a 6-year period but again, this finding is controversial. A recent study of 600 hypertensive T2DM patients with normo- or microalbuminuria, 90 with HF at baseline, showed a faster decline in renal function and progression of albuminuric status over a mean follow-up period of 4 years in patients with persistent HF versus those with normal GFR at baseline and those who had their HF at baseline ameliorated by intensive BP and metabolic control [26]. Conversely, a recent pooled analysis of hyperfiltration studies in T1DM did not show a relationship between HF and increases in AER [27].

In summary, the role of HF as a causative factor in renal disease progression is still unclear and warrants further prospective longer term studies. Whether reversal of HF protects against the risk of development of albuminuria and renal impairment is also worth investigating.

---

**Figure 62.2** Clinical patterns of DKD. A: normo-hypofiltration; B: normo-hyper-normo-hypofiltration; C: normo-hyper-normofiltration; D: persistent normo-filtration.
Progression of DKD to ESRD

Over the past 30 years, a gradual increase in the prevalence of diabetes-related ESRD has been observed, with DKD now accounting for 30–50% of ESRD cases worldwide. Although the absolute number of subjects reaching ESRD continues to increase year by year, an age-specific analysis of ESRD due to diabetes in the United States from 1990–2006 showed that ESRD incidence per 100,000 of the known diabetic population has decreased since 2000 [28]. The incidence of ESRD in T1DM is also declining, based on studies from the USA and Finland. According to the US Renal Data System, in people aged <45 years, the incidence of diabetes-related ESRD decreased by 4.3% per year from 1999 to 2006.

It is unclear to what extent this decrease in ESRD incidence reflects better treatment of early DKD (numerator), including aggressive treatment of glycaemia, hypertension, and dyslipidaemia, as opposed to a higher rate of diagnosis of T2DM during the same interval (denominator). By contrast, in urbanized parts of Asia as well as in socioeconomically deprived parts of the world, the incidence of ESRD due to diabetes continues to increase [29].

Alteration of the rate of progression of DKD to ESRD requires early and intensive management of several risk factors. The most important of these is control of BP, in association with control of glycaemia, dyslipidaemia, smoking cessation, and avoidance of acute kidney injury associated with dehydration or exposure to radiocontrast agents. In hypertensive patients with T2DM and microalbuminuria this so-called multifactorial intervention nearly halved total mortality and ESRD in the Steno-2 Study [15].

While assessing incidence rates of ESRD, it is important to consider the competing risks of renal and cardiovascular endpoints in patients with early and late DN. A meta-analysis of all-cause mortality in patients with CKD (including DKD) showed that two parameters of renal involvement—albuminuria ≥1.1 mg mmol⁻¹ and/or renal impairment with eGFR <60 mL min⁻¹ per 1.73m²—predicted all-cause and cardiovascular mortality. This confirms the data from the UKPDS showing a similar risk of progression though the stages of albuminuria as cardiovascular death. It follows that any intervention which decreases cardiovascular deaths, such as statin therapy, will paradoxically increase numbers of subjects who are at risk of progressing to ESRD.

Gene variants in DKD

Ethnic and familial variability in the risk of CKD in diabetes raises the possibility that genetic factors impact upon the role of the three dominant clinical risk factors for DKD: duration of diabetes, glycemic control, and BP control. However, despite a large number of candidate genes being proposed for predicting DKD, the contribution of individual gene variants to the risk of DKD has achieved relatively modest odds ratios of 1.05–1.5 [30]. Furthermore, in contrast to genetic studies of nondiabetic CKD in the African-American population, findings of genetic variants conferring increased susceptibility have varied in studies of DKD. This inconsistency of results may be explained by inadequate sample size, varying ethnicity of participants, and differences in the duration of diabetes in the DKD group compared with controls.

Another variable complicating the interpretation of genetic studies is the phenotype of DKD. In early DKD, the development of proteinuria on the one hand, and the onset of a decline in GFR on the other, may occur at different rates [31] and may be influenced by different genetic factors [32]. Hence, studies in which the definition of DKD is according to proteinuric status will produce different results from studies where DKD is defined according to GFR trajectory or attained GFR.

A recent meta-analysis of genetic data from 3162 case subjects and 3845 controls from three studies of patients with T1DM and DKD, defined by ESRD, failed to replicate most of the previously reported genetic associations for DKD [33]. This research design targeted specific genes and therefore avoided statistical power issues in previous studies. Larger and more comprehensive studies are therefore needed, firstly to allow prediction of DKD at the time of diagnosis of diabetes, and secondly to facilitate the development of new approaches to therapy of early DKD before it becomes irreversible.

In summary, linkage and association studies in DKD have identified several chromosomal regions and candidate genes that explain a relatively small fraction of the total genetic risk associated with DKD. So far, this fraction does not compare with the advances made in understanding the genetic basis of nondiabetic renal disease in African Americans.

A novel area of gene related research involves the quantification of circulating and urinary micro RNAs. Micro RNAs are small, noncoding transcripts which influence posttranscriptional degradation of messenger RNA [34]. Circulating micro RNAs are disease biomarkers which have potential for the detection and monitoring of DKD.

Histologic changes of DN

Glomerular changes

Glomerular structure is normal in patients with T1DM at the time of diagnosis [35], indicating that hyperglycemia is necessary for the development of DKD. However, the key changes of DN, glomerular basement membrane thickening and mesangial expansion, may be evident within 2 and 5 years, respectively, of T1DM diagnosis. Diffuse and nodular mesangial expansion is an important marker of functional decline in DN. Studies in T1DM have shown a strong inverse correlation between mesangial volume fraction and GFR, probably because the expanding mesangium causes a reduction in the filtration surface area of contiguous glomerular capillaries [36]. While it is generally
accepted that the mesangial expansion and thickening of the glomerular basement membrane characteristic of DN correlates with the degree of proteinuria, this has not been a universal finding, with background RAS inhibition therapy likely playing a confounding role. Of note, the glomerular changes of DN can be reversed when assessed 10 years (but not 5 years) after pancreas transplantation.

Hyaline changes can involve both the afferent and efferent glomerular arterioles. The replacement of glomerular arteriolar smooth muscle cells by hyaline material is associated with the extent of global glomerulosclerosis, suggesting that glomerulosclerosis in DN may derive in part from vascular pathology.

Tubulointerstitial changes

Early in the natural history of DKD, the changes of tubulointerstitial disease are subtle and may be overlooked [37]. However, DKD of long duration is associated with prominent tubulointerstitial pathology, including interstitial fibrosis, tubular atrophy, and mononuclear cell infiltration. The process is related to vascular changes and may be secondary to chronic ischemia. In contrast to patients with T1DM, renal biopsies in patients with T2DM and low degrees of proteinuria may show atubular glomeruli and atrophic tubules, collectively termed glomerulotubular dysjunction. The presence of glomerulotubular dysjunction in DKD may hasten the decline in GFR.

DKD in the absence of elevated AER

An increase in AER classically heralds incipient DKD. However, studies in normotensive patients with T1DM who also have a normal AER and GFR have demonstrated mesangial volume fractions ranging from normal to levels bordering on those usually associated with overt DN. Thus, it now appears that a normal AER does not preclude the existence of structural abnormalities, while the converse also appears true.

The structural basis of DKD without elevations in AER in T2DM remains to be elucidated. However, in patients with T2DM, microalbuminuria and preserved GFR (mean 101 mL min⁻¹ per 1.73m²) findings on renal biopsy are heterogeneous. They include normal or near-normal glomerular and tubulo-interstitial appearance, typical DN with emphasis on glomerular mesangial expansion and basement membrane thickening, or an atypical histologic picture comprising disproportionately severe interstitial/tubular/vascular damage and relatively preserved glomeruli.

Non-DKD in people with diabetes

In patients with T1DM and overt renal disease, the prevalence of nondiabetic CKD is uncommon (<5%). In T2DM, by contrast, up to 25% of patients with CKD have been reported to have nondiabetic CKD. However, this may reflect a clinical bias in the selection of patients for renal biopsy.

In summary, the histologic changes underlying renal dysfunction are much more heterogeneous in patients with T2DM than in patients with T1DM, although patients with T2DM and diabetic retinopathy have an increased likelihood of developing renal pathology similar to that in T1DM. The relationships between the natural history of CKD and the contributions of global glomerulosclerosis, arteriolar hyalinosis, and chronic tubulo-interstitial injury remain inadequately described in people with T2DM.

Prevention, screening, and monitoring of DKD

DKD does not occur in the absence of hyperglycemia and, hence, preventing the transition from normal glucose tolerance to diabetes is the ultimate means of preventing the development of this complication. Studies investigating the prevention of DKD have mainly focused on the development of microalbuminuria. However, currently attention is also being placed on the concept of trying to prevent an early decline in GFR (before reaching 60 mL min⁻¹ per 1.73m²), although data are sparse on how best to achieve this. In T2DM, very good BP control appears to be an important factor in preventing the development of microalbuminuria. Recent studies have also investigated the role of RAS blocking agents in preventing the development of microalbuminuria.

In patients with T2DM and hypertension, angiotensin converting enzyme (ACE) inhibitors attenuate the progression from normoalbuminuria to microalbuminuria. However, in normotensive patients with T1DM, the angiotensin II receptor blocker (ARB) losartan, or the ACE inhibitor enalapril, did not reduce the 5-year cumulative incidence of microalbuminuria compared to placebo treatment. Furthermore, the use of RAS blocking agents did not slow the development of early renal structural characteristics of DN. Possibly, there has to be sufficient intrarenal RAS activation, for example in hypertensive patients or those with late nephropathy, before the benefits of RAS inhibition become apparent.

Screening patients at risk for DKD usually involves testing for the presence of persistent microalbuminuria and estimating GFR. Laboratory-based screening for microalbuminuria can be achieved in a number of ways:

- a 24-hour urine collection for the estimation of AER, accepted as the gold standard with the added advantage of allowing for the estimation of creatinine clearance;
- a spot urine collection for the estimation of the albumin to creatinine ratio (ACR);
- a timed 4-hour or overnight collection for the estimation of AER.

Alternatively, screening for microalbuminuria can be performed on spot or early morning urine samples using special
reagent strips. The reference ranges of albuminuria for AER and ACR 24-hour urine collections and spot urine samples for ACR are described in Table 62.1. At least two out of three consecutive estimations of albuminuria should fall into the microalbuminuric range before a diagnosis of persistent microalbuminuria is made. Increases in albuminuria into the microalbuminuric range may occur transiently with exercise, urinary tract infection, uncontrolled hyperglycemia, and cardiac failure. In patients with T1DM, it is recommended that annual screening for microalbuminuria should commence approximately 5 years after diagnosis, as the development of diabetic complications is rare before this time. In contrast, given that approximately 20% of newly diagnosed subjects with T2DM have microalbuminuria, it is recommended that all subjects with T2DM are screened from the time of diagnosis.

Current gold standard methods for determining GFR, employing the clearance of radioisotopes or nonradionlabeled markers, are time consuming and expensive, and thus not easily adaptable to routine clinical practice. Therefore, creatinine has been used as an endogenous marker of GFR for many decades. Unfortunately, the influence of nonrenal factors on serum creatinine concentrations, including age, gender, ethnicity, muscle mass, and dietary protein intake, limit its usefulness as a marker of GFR. To overcome the limitations of using creatinine alone, there are now equations to estimate GFR based on serum creatinine as well as age, sex, and race, namely, the Modification of Diet in Renal Disease (MDRD) equation [38].

In patients with GFR in the normal or hyperfiltration range, GFR is grossly underestimated by the MDRD formula by approximately 10–40 mL min⁻¹ per 1.73 m². To overcome the limitations of the MDRD equation a new creatinine-based formula, based on the original variables used in the MDRD formula, has been devised to improve estimation of GFR in individuals with a GFR >60 mL min⁻¹ per 1.73 m², that is, the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation [39]. However, the CKD-EPI equation has yet to be rigorously tested in subjects with diabetes with high-normal GFRs. In contrast, recent studies have suggested that estimates of GFR based on serum cystatin C concentrations provide a simple and accurate method for detecting and monitoring an early decline in renal function. Although cystatin C is well established as a research tool, it has yet to make the transition to use in routine clinical practice.

Once the urinary albumin threshold for microalbuminuria has been reached, AER or ACR should be measured every 3–6 months. Measurement of serum creatinine to allow for automatic reporting of an eGFR should also be made every 3–6 months. However, as discussed earlier, existing methods for estimating GFR are only robust below a threshold of 60 mL min⁻¹ per 1.73 m². The implication is that approximately one half of total kidney filtration rate can be lost before it is possible to accurately detect a decline in renal function using current clinical parameters. Nevertheless, serial measurements of eGFR between 90–60 mL min⁻¹ per 1.73 m² are useful for detecting trends in GFR trajectory which may occur without changes in AER.

### Management of DKD

#### Glucose control (Table 62.4(a))

Glycemia is the major determinant for the onset of DKD. Both the DCCT in T1DM and the UKPDS in T2DM demonstrated the strong relationship between glycemic control and the risk of the development of diabetic microvascular complications, without a clear-cut threshold of HbA1c concentration. However, a recent sophisticated observational analysis from the ADVANCE study involving subjects with T2DM suggested that within the range of HbA1c studied (5.5–10.5%), below an HbA1c threshold of 6.5% there was no significant increase in the risks of retinopathy or late renal complications including macroalbuminuria, doubling of serum creatinine, the need for renal-replacement therapy, or death due to CKD [40].

<table>
<thead>
<tr>
<th>eGFR (mL min⁻¹ per 1.73 m²)</th>
<th>T1DM</th>
<th>T2DM</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥60</td>
<td>Attained A1c 7.2% over 6.5 yrs (DCCT) (i) reduces incident micro- and macro-AER (UKPDS 33) [64]. (ii) Reduces GFR decline to &lt;60 from 5.5% (standard) to 2.0% (intensified) over 22 yrs DeBoer [63] (DCCT/EDIC)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(i) Attained A1c 7.0% over 10 yrs reduces incident micro- and macro-AER (UKPDS 33) [64]. (ii) Attained A1c 6–7%: reduces incident micro- AER, with no effect on GFR decline (ADVANCE), but may reduce progression to ESRD in macro- AER (NNT 41) [42]. (iii) Combined glucose and BP control reduces diabetes-related events more than either intervention alone (UKPDS 75) [65].</td>
<td></td>
</tr>
</tbody>
</table>
In the DCCT, the development of microalbuminuria in T1DM patients was reduced by 34% and macroalbuminuria by 56% through tight glycemic control (HbA1c < 7.0%). An interventional analysis of the UKPDS in T2DM patients demonstrated that reducing mean HbA1c concentrations from 7.9 to 7.0% was associated with an absolute risk reduction for the development of overt nephropathy of 11% over 12 years, and a relative risk reduction of 30% for the transition from normo- to microalbuminuria.

More recently, results from the DCCT/EDIC research group have shown that in T1DM, the long-term risk for the development of an impaired GFR (\( < 60 \text{mL min}^{-1} \text{per 1.73m}^2 \)) was significantly lower for people who were initially treated to achieve tight glycemic control. In T2DM patients, a meta-analysis of intensive glucose control studies has shown that compared to conventional control, intensive glycemic control reduces the risk of developing microalbuminuria (risk ratio 0.86; 95% CI 0.76–0.96) and macroalbuminuria (risk ratio 0.74; 95% CI 0.65–0.85). However, the risk for the development of ESRD was not significantly different compared with conventional glucose control. This result was possibly related to the relatively low incidence of ESRD (<1.5%) compared with that of microalbuminuria (23%) and macroalbuminuria (5%) [41].

By contrast, a recent analysis of the ADVANCE study has shown that intensive glucose control not only reduces the incidence of micro- and macroalbuminuria but may also reduce the incidence of ESRD to a lesser degree. One ESRD event was prevented when 41 participants with macroalbuminuria at baseline were treated intensively over 5 years.

In summary, the HbA1c threshold for the development of DKD remains to be clearly defined but is likely to be around 6.5%, the same as the currently accepted cut-off for the diagnosis of diabetes. Recent evidence has shown that intensive glucose control not only reduces the incidence of micro- and macroalbuminuria but may also reduce the incidence of ESRD to a lesser degree. In this setting, there is increasing clinical awareness that appropriate glycemic targets to prevent diabetes-related complications need to be individualized for each patient.

**Blood pressure (BP) control (Table 62.4(b))**

In patients with T1DM, elevations in BP are usually due to underlying DKD as, typically, BP only begins to rise around the time of transition from micro- to macroalbuminuria. In contrast, hypertension may be present in approximately one third of patients with T2DM at the time of diagnosis. In this setting, the etiology of hypertension is most likely multifactorial and represents a component of the metabolic syndrome. Regardless of the sequence or underlying causes of hypertension in T1DM and T2DM, many studies have demonstrated that high BP accelerates the progression of DKD and that aggressive BP lowering retards the deterioration in renal function.

The recommended BP target for people with diabetes is 140/80 mmHg [43]. If persistent proteinuria is present, then a more stringent target of less than 130/75 mmHg has been recommended, although this is not achievable in most elderly subjects with T2DM. As discussed later, antihypertensive
medications that interrupt the RAS appear to have renal protective effects over and above those expected from BP lowering effects alone. However, this superior effect is progressively reduced as blood pressure declines. Achieving BP goals commonly requires the combination of three to four different antihypertensive medications, particularly in subjects with T2DM. People with diabetes, especially those with DKD, are also at higher risk of CV disease; therefore, achieving BP targets offers both renal and CV disease protection. It is recommended that initial antihypertensive therapy be commenced with an agent that inhibits the RAS. While diuretics and dihydropyridine calcium channel blockers are useful additional agents, the ACCOMPLISH [44] trial suggests that adding a calcium channel blocker to an ACE inhibitor offers advantages over the addition of a thiazide diuretic. The choice of antihypertensive agents will depend to some extent on the subject's clinical circumstances, as these may preclude certain drug combinations (e.g. the combination of β-blockers and nondihydropyridine calcium channel blockers), or dictate the early introduction of specific antihypertensive agents.

Although observational studies suggest a continuous relationship between BP levels and renal and CV events in people with diabetes, very few interventional studies have achieved BP levels below 130/80 mmHg. The results of both the ACCORD-BP [45] trial and INternational VErapamil SR Trandolapril Study (INVEST) [46], which have attained BP levels of <130/80 mmHg, suggest that lowering BP below these levels has little impact on reducing clinical events below those seen with moderate BP control. Ambulatory blood pressure monitoring (ABPM) provides measurements of several variables that are of potential value in estimating the risk of clinical outcomes in diabetes. These include an estimate of “true” BP levels and the identification of the “white-coat” effect as well as the measurement of the diurnal rhythm of BP that allows patients to be classified into nocturnal “dippers” and “nondippers” (a difference between mean daytime and nocturnal BP of <10%). Nondipping can be treated by administering BP-lowering drugs in the evening. Another use of ABPM is the recording of random variability of BP and ambulatory heart rate as measures of autonomic function. Importantly, it should be noted that the clinical trial information presented earlier, especially relating to BP targets, has been based on clinical BP recordings.

Renin-angiotensin system (RAS) inhibition (Table 62.4(b))

In patients with diabetes, activation of the intrarenal RAS produces hemodynamic and nonhemodynamic effects that contribute to the development and progression of DKD. There is now good evidence suggesting that interrupting the RAS with an ACE inhibitor or an ARB results in renal and CV protective effects over and above those observed by BP lowering alone. In a landmark trial, Lewis and colleagues [47] demonstrated that the ACE inhibitor captopril reduced the risk of achieving the combined endpoint of death, dialysis, and transplantation compared to BP control alone achieved with non-RAS blocking agents in subjects with T1DM and advanced nephropathy (macroalbuminuria). In another important study, captopril significantly reduced the progression from micro- to macroalbuminuria and prevented an increase in AER in patients with T1DM [48]. A subsequent systematic review of the effects of ACE inhibition in patients traditionally classified as normotensive, who also had microalbuminuria, has been performed (ACE Inhibitors Trialist Group 2001). It concluded that ACE inhibition use significantly reduced the progression to macroalbuminuria and increased the chances of remission to normoalbuminuria. However, the Renin-Angiotensin System Study (RASS) [49], using angiotensin receptor blockade or ACE inhibition in normotensive, normoalbuminuric patients with T1DM did not show any effects on the rate of progression of AER from normo- to microalbuminuria [50].

For patients with T2DM and microalbuminuria, the use of either ACE inhibitors or ARBs has been shown to reduce urinary albumin excretion and CV events, but has not been shown to affect renal function decline [53]. By contrast in patients with T2DM, hypertension, and macroalbuminuria, the use of an ARB, superimposed on background antihypertensive therapy, slows the progression to ESRD compared to placebo (IDNT [51] and RENAAL [52]) and the decrease in albumin excretion rate predicts the subsequent rate of decline of renal function [53].

Dual blockade of the RAS with an ACE inhibitor and ARB in subjects with T2DM and microalbuminuria has been demonstrated to be more effective in reducing BP and decreasing albuminuria than either agent as monotherapy. However, outcome studies have highlighted an increase in adverse outcomes with this approach (ONTARGET study) [54]. The combination of an ACE inhibitor and ARB is therefore generally not advised but can still be considered as a therapeutic option in patients with progressive albuminuria in the macroalbuminuric range or in patients with refractory hypertension (>160/100 mmHg).

Dyslipidemia (Table 62.4(c))

The role of dyslipidemia in the initiation and progression of DKD remains poorly defined. Several cross-sectional and prospective studies have shown associations between dyslipoproteinemia, specifically elevation in apoB-100 containing lipoproteins and low HDL-cholesterol, and albuminuria in subjects with diabetes.

In a post hoc analysis from the ADVANCE study in T2DM, low baseline HDL concentrations were associated with a significantly higher risk for the development of new microalbuminuria and macroalbuminuria.

Studies involving the FinnDiane cohort of T1DM patients have suggested that higher triglyceride concentrations are associated with progressive albuminuria, whereas total cholesterol concentration was the only lipid parameter associated with progression to ESRD in an adjusted Cox regression model that
Table 62.4(c) Effects of interventions on renal outcomes in diabetes

<table>
<thead>
<tr>
<th>Lipid control:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>eGFR (mL min⁻¹ per 1.73m²)</td>
<td>T1DM</td>
</tr>
<tr>
<td>≥60</td>
<td>No large studies have assessed statin effects on AER or GFR in T1DM</td>
</tr>
<tr>
<td></td>
<td>(i) Compared with placebo, fenofibrate reduces AER and slows annual eGFR decline from 6.9 to 1.9 mL min⁻¹ per 1.73m² [57].</td>
</tr>
<tr>
<td></td>
<td>(ii) Statins reduce CV events but do not affect DN [70].</td>
</tr>
<tr>
<td></td>
<td>(iii) Fenofibrate reduces CV events only in subgroup with ↑ TG/↓ HDL [71].</td>
</tr>
</tbody>
</table>

Lifestyle factors

Obesity per se is associated with the development of a specific form of focal and segmental glomerulosclerosis, characterized by massive proteinuria. The underlying pathophysiology includes insulin resistance, hypertension, adiponectin deficiency, and hyperaldosteronism, which could all contribute to the development of proteinuria in obese people with diabetes. There is some evidence that weight loss reduces proteinuria and stabilizes progression of CKD in obese people with and without diabetes. In obese people with DKD, weight loss should be encouraged and promoted through the use of diets low in calories and saturated fats. Weight loss achieved with bariatric surgery has been associated with significant reduction in albuminuria.

Several observational studies have documented an association between smoking and DKD. In particular, smoking appears to promote the initiation of microalbuminuria and the subsequent transition to macroalbuminuria. Furthermore, evidence exists to suggest that cessation of smoking retards the progression of DKD. Thus, despite a lack of definitive interventional studies for smoking, there is a strong rationale for the inclusion of smoking cessation in the management of subjects with DKD.

Other interventions

While the American Diabetes Association guidelines for the management of DKD recommend a protein intake of 0.8 g kg⁻¹ d⁻¹ in patients with overt nephropathy, the difficulty experienced by patients in following a low-protein diet long term should be recognized.

Anemia is common in patients with diabetes and has been considered a potential risk factor for the progression of DKD [59]. However, evidence has emerged from randomized controlled trials that correcting anemia with erythropoiesis-stimulating agents in subjects with DKD prior to dialysis is associated with an increased risk of serious adverse outcomes and no evidence of any beneficial effect. Therefore, treatment of anemia with erythropoiesis-stimulating agents in subjects with
diabetes and predialysis CKD is generally not recommended. Patients should still be screened for other reversible causes of anemia, such as iron deficiency and treated appropriately. Recent studies do not support the universal use of aspirin for the prevention of CV events in patients with diabetes. However, given the increased CV risk associated with CKD it is reasonable to consider aspirin for the primary prevention of CV disease in people with DKD. Clopidogrel does not appear to be a suitable alternative antiplatelet drug for people with DKD. Newer antiplatelet drugs have become available in recent years, but their effectiveness and safety in lowering CV risk in diabetes requires further evaluation.

In general, patients with diabetes and a GFR <30 mL min⁻¹ per 1.73m² should be referred to a nephrologist in preparation for the commencement of renal replacement therapy. Referral should also be considered for patients found to have nephrotic range proteinuria, or when the diagnosis of the underlying cause for renal impairment is uncertain.

**Multifactorial interventions**
The management of people with DKD involves strict attention to the common risk factors for kidney and CVD disease; namely, hyperglycemia, hypertension, dyslipidemia, and smoking. Intensive multifactorial intervention was first performed in the Steno-2 study [15] which used treatment targets of HbA1c <6.5%, systolic BP <130 mmHg, diastolic BP <80 mmHg, fasting cholesterol <4.5 mmol L⁻¹, and fasting triglycerides <1.7 mmol L⁻¹. The study involved the universal use of RAS inhibitors, aspirin, the aggressive use of statins, and attention to lifestyle factors. This approach decreased the risk of developing overt proteinuria by 60% and CV events by approximately 50% in people with T2DM, hypertension, and microalbuminuria followed for 8 years. These results were achieved despite only a minority of subjects reaching the targets for glycemia and BP control. In addition, a 5-year observational study that monitored the development of complications after the completion of the interventional phase of the Steno-2 study showed that the risk of developing ESRD was significantly reduced in people assigned to the above multifactorial, target-driven management strategy compared to those who originally received conventional treatment. The effects of intensified glucose, BP and lipid interventions on renal and CV outcomes in people with diabetes are summarized in Table 62.4(a).

**Novel interventions**
Despite the wide use of available therapies that target the RAS, people with DKD still progress to ESRD. Current strategies based on RAS inhibition are only partially effective. This may, in part, be due to the interference of ACE inhibitors or ARBs with negative feedback mechanisms resulting in a reactive increase in plasma renin activity. Aliskiren, an oral direct renin inhibitor, was investigated in the Evaluation of Proteinuria in Diabetes (AVOID) trial and found to reduce ACR in subjects with T2DM and overt DN who were already receiving the maximal recommended doses of an ACE inhibitor. However, a recent clinical endpoint study (The Aliskiren Trial in T2DM Using Cardio-Renal Endpoints (ALITTUDE)) was ceased due to adverse outcomes in participants randomized to the combination of aliskiren plus another RAS-blocking agent. It remains unclear whether the adverse outcomes in the aliskiren group were related to the degree of BP-lowering or to some other mechanism.

Unfortunately, the potential benefits of promising DKD interventions such as aminoguanidine, sulodexide, and bardoxolone have not come to fruition when tested in large clinical trials. Recently, paricalcitol, a vitamin D receptor activator, has been shown to reduce albuminuria in subjects with T2DM and macroalbuminuria without an increased incidence of hypercalcemia or other serious adverse events. Whether vitamin D activation treatment results in a preservation of GFR has yet to be tested in clinical trials. Other vitamins, such as vitamin B complexes, have no effect on proteinuria but have caused a significant decline in GFR compared with placebo-treated subjects. Benfotiamine also does not appear to have any effect on urinary albumin excretion in subjects with T2DM. Other agents, such as the endothelin antagonists, reduce albuminuria but currently have an unacceptable side effect profile. It is possible that even older drugs, such as allopurinol, may be shown to play an important role in ameliorating the progression of DKD in future years, with raised uric acid concentrations emerging as an important risk factor for progressive renal disease. Lastly, preclinical studies of DKD suggest that specific antifibrotic agents may provide another novel approach to slowing the progression of DKD. The antifibrotic agent, pirfenidone, has been shown to slow the rate of decline of estimated GFR over 12 months in subjects with established diabetic nephropathy [60]. However, this requires confirmation in a longer study before pirfenidone can be considered for clinical use.

**References**
6. Coughlan MT, Thorburn DR, Penfold SA, et al.: RAGE-induced cytosolic ROS promote mitochondrial superoxide generation...


30 Brorsson C, Pociot F: Genetics of diabetic nephropathy in diverse ethnic groups. *Contributions to Nephrology* 2011;170:8–18.


39 Stevens LA, Schmid CH, Greene T, et al.: Comparative performance of the CKD Epidemiology Collaboration (CKD-EPI) and the Modification of Diet in Renal Disease (MDRD) Study equations for estimating GFR levels above 60 mL/min/1.73 m². *American Journal of Kidney Diseases* 2010;56(3):486–495.


CHAPTER 63
Pathology of human diabetic neuropathy

M. Mojaddidi1, A.K. Sharma2 and Rayaz A. Malik3
1Taibah University, Al-Madinah Al-Munawarah, Saudi Arabia
2Uttaranchal Ayurvedic Medical College, Dehradun, India
3Weill Cornell Medical College, Doha, Qatar

Key points
- Diabetic neuropathy has multiple manifestations, which result in increased morbidity and mortality.
- Demyelination and axonal degeneration are the primary pathologic features of diabetic neuropathy.
- Skin biopsy has enabled early identification of small fiber pathology in diabetic neuropathy.
- Corneal confocal microscopy represents a novel, noninvasive clinical technique for the early diagnosis, assessment of progression and benefits of therapeutic intervention, making it an ideal surrogate endpoint for diabetic neuropathy.

Introduction
Detailed study of the pathology of diabetic neuropathy requires the quantification of structural alterations in nerve or nerve-containing tissue from patients with diabetes. Biopsy of the sural nerve represents the gold standard for accurate morphometric evaluation of neuropathology in patients with peripheral neuropathy. It allows the evaluation of large myelinated, and in particular small unmyelinated nerve fibers as well as blood vessels. Inherent with the focus on neurophysiologic abnormalities in diabetic neuropathy, large myelinated nerve fibers have been studied the most. Until recently very few studies have assessed small fibers, even though small fiber/autonomic dysfunction has been shown to have comparable sensitivity/specificity for detecting neuropathy [1] and they constitute 79.6% [2] to 91.4% [3] of peripheral nerve fibers. Moreover, they underlie painful diabetic neuropathy and are key to the genesis of foot ulceration via sudomotor dysfunction [4], reduced pressure-induced vasodilatation [5–7], and, of course, impaired heat and pain perception [8]. This chapter will consider the significant contribution that sural nerve biopsy studies have made to our understanding of the pathology and pathogenesis of human diabetic neuropathy and will then also evaluate the increasing contribution of recent studies utilizing skin biopsies [9] and the novel technique of corneal confocal microscopy in evaluating human diabetic neuropathy [10–12].

Nerve biopsy
The majority of nerve biopsy studies have been undertaken in the sural nerve due to the fact that it is distal, sensory, and relatively easy to biopsy [13]. In the upper limb the posterior interosseous nerve, a sensory branch of the radial nerve, has also been biopsied and studied [14]. Morphometric evaluation has involved predominantly light and electronmicroscopy in glutaraldehyde fixed tissue, although some studies in amyloid neuropathy [15,16] and diabetic lumbosacral radiculoplexus neuropathy [17,18] have undertaken immunohistologic evaluation in paraffin sections to provide considerable insight into the pathogenesis of these conditions. Whilst sural nerve biopsy is a valuable tool for the study of diabetic neuropathy, concerns of persisting pain and numbness [19,20] from traditional whole sural nerve biopsy have led to the use of fascicular biopsy techniques [21].

Myelinated fibers
Density
A loss of myelinated fibers is the most striking feature of diabetic neuropathy and can be assessed by quantifying myelinated fiber density (MNFD) which is derived by counting the number of myelinated fibers within each fascicle and expressing the result as no. mm−2. Whilst most studies consistently demonstrate a progressive reduction in myelinated fiber density with increasing severity of neuropathy (Figure 63.1(a–d)), it is important to note that loss of myelinated fibers is not the earliest feature of diabetic neuropathy and indeed demyelination precedes a loss
of myelinated fibers [3]. Of course the loss of myelinated fibers can be extreme in type 1 diabetic patients with somatic and autonomic neuropathy, such that in a group of five patients aged 22–34 with an uncommonly severe symmetric polyneuropathy, myelinated fiber density was reduced to 20 +/- 14% of that in five control subjects [22,23]. Similarly, in diabetic patients with more advanced neuropathy and concomitant autonomic deficits the extent of myelinated fiber loss is extremely pronounced [22,24–27]. Myelinated nerve fiber loss has been related to electrophysiologic measures of neuropathic severity [28–30] and elevated vibration perception [22]. There are precise relationships between different structural components of the myelinated fibers, thus as early as 1966 Lascelles and Thomas demonstrated a linear relationship between internodal length and axonal diameter [31]. The seminal study which defined a relationship between structure and function involved the morphologic study of sural nerves in five patients with mild sensory involvement, four with severe, symmetrical sensory-motor polyneuropathy, and three with multiple mononeuropathy [29]. All demonstrated a loss of large and small myelinated and unmyelinated fibers and teased fiber studies showed predominant segmental remyelination. The relation between the conduction velocity and the diameter of the largest fibers indicated that approximately 20–30% of the conduction slowing was due to causes other than fiber loss [29], such as changes in internodal length [32] and, of course, demyelination which was first reported in diabetic neuropathy by Thomas and Lascelles in 1965 [33]. In a sural nerve biopsy study MNFD
was reduced compared with control biopsies (4042+/−2090 (+/− SD) vs. 6800+/−1100 mm²; p < 0.01) and correlated significantly with sural sensory conduction velocity (r = 0.84; p < 0.001), sural action potential amplitude (r = 0.74; p < 0.001), peroneal motor conduction velocity (r = 0.58; p < 0.02), and median sensory amplitude (r = 0.64; p < 0.01) but did not correlate with any quantitative sensory test [34]. Vibration sense threshold has been positively correlated with the total number of myelinated fibers [22]. In a longitudinal (7−10-year) follow-up study of 10 subjects with T2DM and 10 subjects with IGT a baseline MNFD (<or = 4700 fibers per mm²) was associated with a decline in peroneal amplitude (p < 0.02) and conduction velocity (p < 0.04), as well as median nerve sensory amplitude (p < 0.05) and motor conduction velocity (p < 0.04) [35]. More recently two studies have defined progression of diabetic neuropathy as a >500 per mm² loss of myelinated fiber density over 12 months and performed clinical and genomic studies to demonstrate a link between hypertriglyceridemia and a range of other putative genes with the progression of neuropathy [36,37]. Such a rapid and extreme reduction in myelinated fiber density does not reflect the natural history of myelinated fiber loss, and likely delineates the rapid progressors; therefore these findings may not necessarily be applicable to all diabetic patients who develop neuropathy. Evaluation of the regenerative cluster density (Figure 63.2) has also been undertaken to attempt to understand the balance between degeneration and regeneration of myelinated fibers [23,38]. In a sural nerve biopsy study the number of regenerating fibers, identified by light microscopy, was found to decline in proportion to the reduction in total myelinated fiber density; however, the relative number of regenerating fibers was significantly greater in patients with T1 compared to T2DM [38]. However, no relationship has been found between the regenerative capacity assessed by quantifying regenerative clusters and painful symptoms of diabetic neuropathy [25]. Apart from a generalized loss of myelinated fibers, a change in the spatial distribution of nerve fiber loss, proximally [39] has been demonstrated to suggest a vascular basis for diabetic neuropathy, although another study has demonstrated comparable patchy fiber loss in a series of sural nerve biopsies from patients with diabetic neuropathy and type 1 hereditary motor and sensory neuropathy, questioning a vascular causation [40]. However, fascicular sural nerve morphometry has demonstrated significant differences between 11 T1DM patients and 17 T2DM patients with comparable duration and severity of hyperglycemia and severity of neuropathy [41]. Thus in patients with T2DM there was evidence of predominantly Wallerian degeneration with the spatial pattern of fiber loss quantified using the intrafascicular coefficient of variation demonstrating a multifocal pattern, consistent with ischemia. Pathology has also been quantified in the posterior interosseous nerve (PIN) in the forearm [14] in 26 diabetic and 20 nondiabetic patients with carpal tunnel syndrome (CTS) and 13 control subjects. PIN myelinated nerve fiber density was significantly reduced in diabetic and nondiabetic patients with CTS compared to control subjects and diabetic patients had a significantly lower density than nondiabetic patients [42].

**Axonal atrophy**

Based on a series of experimental studies showing a reduction in the distal myelinated fiber axonal caliber in relation to the number of myelin lamellae compared to controls it has been suggested that axonal atrophy occurs in diabetic neuropathy [43] (Figure 63.3(a,b)). In a study of 15 diabetic patients with mild neuropathy and 14 control subjects, whilst the diabetic patients demonstrated a significant reduction in myelinated fiber density, there was no change in fiber and axonal area, or g-ratio, compared to control subjects [30]. A detailed electron-microscopic morphometric study of sural nerve biopsies from 8 patients without and 25 patients with diabetic neuropathy and 24 healthy subjects undertook a more detailed assessment of axonal atrophy and demonstrated that the regression lines of the natural log (In) of axonal area and index of circularity (IC) (an index that is decreased with atrophy or shrinkage) on the number of myelin lamellae and the rate of adaxonal sequestration did not differ significantly in large myelinated fibers of diabetic patients compared to controls. However, interestingly, significant differences were found in regression lines of both In axonal area and IC on the number of lamellae for small myelinated fibers and all fibers, but were
attributed to remyelination and axonal regeneration [44].

Furthermore, in a detailed neuropathologic study of sural nerve biopsies from 6 control subjects and 16 age-matched diabetic patients with different syndromes of sensory polyneuropathy (6 with chronic painful neuropathy [CPN], 4 with newly presenting painful neuropathy [NPN], and 6 with painless neuropathy associated with recurrent neurotrophic foot ulcers [RFU]), fibers with disproportionately large Schwann cells (cytoplasm and myelin) relative to axon caliber were exclusively found in patients with neuropathic pain, suggesting that axonal atrophy may be involved in the generation of neuropathic pain [25]. In a 12-month randomized, placebo-controlled, double-blind clinical trial, sural nerve biopsies were obtained in 10 neuropathic diabetic patients treated with the aldose reductase inhibitor sorbinil and in 6 patients treated with placebo at baseline and follow-up. Axonal atrophy was assessed using three independent morphometric techniques and exhibited a significant recovery in the sorbinil-treated patients [45]. Therefore larger studies are required to confirm or refute the occurrence of axonal atrophy as its occurrence, or lack of, may have significant connotations towards our understanding the mechanistic basis and perhaps treatment of diabetic neuropathy, especially as recent findings in other neuropathies have begun to unravel the molecular basis for this change [46,47].

**Axoglial dysjunction**

In 12 diabetic patients with minimal neuropathy, whilst myelinated fiber density, fiber and axonal area, and g-ratio were normal, teased fiber studies showed significant paranodal abnormalities, segmental demyelination and remyelination without axonal degeneration [3]. Therefore paranodal abnormalities, in the form of “axo-glial dysjunction” possibly suggesting an early Schwann cell abnormality, may be important in the genesis of diabetic neuropathy. Indeed this term was first coined in 1986 when morphometric analysis in long-term spontaneously diabetic BB-Wistar rats demonstrated "a striking disappearance of paranodal axo-glial junctional complexes that was not corrected by insulin replacement" [48]. It was hypothesized that loss of these junctional complexes, may allow lateral migration of axolemmal Na channels away from nodes of Ranvier, leading to diminished nodal Na permeability and resultant nodal conduction delay. Morphometric assessment of this abnormality is of course very much dependent on adequate fixation and therefore not reliable. The authors who originally described this abnormality have demonstrated a significant improvement in axo-glial dysjunction after a 12-month randomized, placebo-controlled, double-blind clinical trial of the aldose reductase inhibitor sorbinil. However, two independent studies of sural nerve biopsies from diabetic patients have failed to confirm the presence of axo-glial dysjunction [49,50] and therefore the validity of axoglial dysjunction in human diabetic neuropathy has been questioned [51]. No abnormality was found in a series of 29 patients with diabetic neuropathy when axoglial dysjunction was assessed by quantifying the paranodal dimensions and by assessing high power EM micrographs for any alteration in the attachment of the terminal loops to the axolemma and change in the desmosome-like junction [50].
**Unmyelinated fibers**

The detailed morphologic assessment of unmyelinated fiber pathology requires electron microscopic morphometry to quantify axon density, size, and the density of unassociated Schwann cell profiles. It appears that unmyelinated fibers may be damaged early but show significant repair even in advanced neuropathy. Thus in a study of diabetic patients with minimal evidence of neuropathy the unassociated Schwann cell profile density (Figure 63.4) \((p < 0.04)\) was increased suggestive of degeneration but was accompanied by an increase in the unmyelinated axon density \((p < 0.001)\) with a reduction in the axon diameter \((p < 0.007)\) and an overall shift of the size frequency distribution confirming the presence of axonal sprouts and hence regeneration (Figure 63.4(a,b)) [3]. Similarly, in a study of 15 diabetic patients with mild neuropathy, the percentage of unassociated Schwann cell profiles \((p < 0.0001)\) and axon density \((p < 0.0008)\) were significantly increased with a significant reduction in unmyelinated axon diameter \((p < 0.001)\) with a shift of the size frequency distribution to the left \((p < 0.02)\). In four young type 1 diabetic patients with severe autonomic dysfunction characterized by postural hypotension, fainting, diarrhea, and Argyll Robertson pupils, sural nerves demonstrated a marked reduction \((6 \pm/- 4%\) of that in the controls) in the density of unmyelinated fibers [23]. Although, in a patient with severe postural hypotension and significant autonomic neuropathy there was still evidence of marked unmyelinated fiber regeneration in the form of axonal sprouts with a 77% increase in axon density and 62% reduction in axon diameter [24]. In a series of patients with T1DM subdivided into those with: (a) severe autonomic neuropathy and accompanying painless sensory neuropathy, (b) severe autonomic neuropathy and chronic painful sensory neuropathy, and (c) chronic or acute painful sensory neuropathy without autonomic neuropathy, the density and median diameter of unmyelinated axons was reduced in all three groups [22] and there was no correlation between the occurrence of pain and active degeneration or regeneration of unmyelinated axons [22]. Sural nerve biopsies from four diabetic patients with acute painful neuropathy and two in remission from this condition revealed evidence of unmyelinated fiber degeneration and regeneration in both groups, but remission from pain was found to be associated with less active unmyelinated fiber regeneration [52].

**Microvessels**

Microvascular disease contributes to the genesis of human diabetic neuropathy. Major pathologic abnormalities of the microvessels have been demonstrated in the majority of sural nerve biopsies from patients with IGT [53] and diabetic neuropathy [3,22,23,54–56]. Microangiopathy is most pronounced in the endoneurial capillaries [57], but also occurs in the transperineurial [58] and epineurial [59] microvessels. The key features of the endoneurial microangiopathy are luminal narrowing [53,60], endothelial cell hyperplasia and hypertrophy [30,55,56], pericyte loss [30,56] and basement membrane thickening [30,54,56] (Figure 63.5(a,b)). Progressive endoneurial microangiopathy has been related to increasing severity of neuropathy defined by reduced nerve conduction velocity and loss of myelinated nerve fibers [54,55].

**Extracellular matrix**

The extracellular matrix comprises the space between the vascular and neural elements of the endoneurium and therefore
Pathology of human diabetic neuropathy

Figure 63.5 Electronmicrographs from a control subject (a) and diabetic patient with neuropathy (b) showing thickening of the basement membrane and reduction in luminal size of the endoneurial capillaries (x3000).

Figure 63.6 Electronmicrograph from a diabetic patient showing a collagen pocket (x12000).

studies were undertaken to characterize matrix changes in relation to neuropathology [61]. Increased endoneurial collagenization (type I and III but not type II) was observed in diabetic patients with a relative increase in type VI collagen around Schwann cells, types IV, V and VI collagen around endoneurial microvessels, and types IV and V collagen in the perineurium. The mean diameter of endoneurial collagen fibrils was increased, although the diameter of the fibrils within the basal laminal tubes surrounding degenerated myelinated fibers and within onion bulbs were of normal endoneurial calibre (Figure 63.6). Immunohistologic studies showed normal expression of laminin by Büngner bands with no difference in endoneurial fibronectin expression [61]. In a study of five sural nerve biopsies from patients with T2DM, advanced glycation end-products (AGE) were localized in the perineurium, endothelial cells, and pericytes of endoneurial microvessels and at the submicroscopic level as focal aggregates in the cytoplasm of endothelial cells, pericytes, axoplasm, and Schwann cells of both myelinated and unmyelinated fibers [62]. More recently in a small sample of five sural nerve biopsies from patients with T2DM, collagen V and VI were significantly increased in the perineurium, whilst type IV collagen was normal. Laminin expression was normal, but tenascin expression was increased, especially around the axon myelin units of diabetic patients [63].

Variants of diabetic neuropathy

When diabetic patients develop neurologic deficits which are not consistent with diabetic sensorimotor polyneuropathy (DSPN) it is important to consider variants such as painless motor predominant neuropathy [64], diabetic lumbosacral radiculoplexus neuropathy (DLRPN), or chronic inflammatory demyelinating polyradiculoneuropathy (CIDP). Detailed pathologic and immunohistologic studies of sural [17,18], superficial peroneal [65], and intermediate cutaneous nerve of...
the thigh (ICNT) biopsies and a recent post-mortem study [66] have provided novel insights into the pathophysiology of these less frequent, but clinically challenging variants of diabetic neuropathy (Figure 63.7).

In 1994, the landmark study by Said et al. changed our understanding of DLRPN [67]. A immunohistologic and pathologic study was undertaken in 10 type 2 diabetic patients with diabetic amyotrophy who underwent biopsy of the ICNT, a sensory branch of the femoral nerve [67]. The density of myelinated and unmyelinated fibers showed a variable decrease, with evidence of both axonal degeneration and demyelination on teased fiber studies. However, this was the first study to reveal ischemic nerve lesions in three, vasculitis and inflammatory infiltration in two, and mild inflammatory infiltration in four patients. A number of studies have followed which confirm an inflammatory basis for diabetic amyotrophy. Paraffin sections of sural nerves from 19 patients with DLRPN demonstrated a greater number of ICAM-1 positive vessels with TNF-α expression in Schwann cells and some macrophages as well as increased NFκ-B immunoreactivity in vessels and endoneurial cells which correlated with the number of empty nerve strands and axonal degeneration in teased fiber studies [68].

A series of 22 diabetic patients with painful multifocal diabetic neuropathy due to root, plexus or nerve trunk involvement underwent biopsy of a recently affected sensory nerve. There was a marked reduction in both myelinated (1340+/− 1070 vs. 8370+/− 706) and unmyelinated (5095+/− 6875) nerve fiber density (per mm²), due primarily to axonal degeneration (34+/− 31%) which was associated with a necrotizing vasculitis of perineurial and endoneurial blood vessels, perivascular mononuclear cell infiltrates, and endoneurial hemorrhage [69].

In a recent study, of an initial cohort of 470 patients, detailed neurologic evaluation and nerve biopsy assessment was undertaken in 23 patients with a slowly progressive, severe, painless, symmetrical, motor predominant (~50% in wheel chairs due to weakness) neuropathy, 33 patients with DLRPN, and 25 patients with CIDP. Examination demonstrated pan-modality sensory loss, autonomic abnormalities, and an axonal neuropathy. The nerves studied by biopsy were not specified, but one would assume that these were sural nerves and they demonstrated a reduction in myelinated fiber density, which was severe in 8/23, multifocal (11/23) and confirmed to be axonal degeneration on teased fiber studies, but with evidence of regeneration in the form of regenerative clusters (16/23). There were associated alterations in the epineurial vessels (23/23), vessel wall inflammation (15/23), and microvasculitis (3/23) (Figure 63.7). In contrast, CIDP biopsies revealed demyelination and onion bulbs without axonal loss or microvasculitis.

Given the possible association and increased incidence of CIDP in diabetes [70], a study of sural nerve biopsies from 10 patients has demonstrated epineurial perivascular T-cell infiltrates and immunoreactivity for matrix metalloproteinase-9 in diabetic patients with CIDP which was absent in patients with DSPN alone [71]. Indeed using a cut-off, of 24%

Figure 63.7 Light microscopic image of a sural nerve from a diabetic patient with DLSPN showing epineurial vessel pathology (×500). (For a color version of this figure, please see color plate section.)

MMP-9-positive cells, CIDP could be detected with a sensitivity of 70% and a specificity of 100% in diabetic patients.

Skin biopsy

Skin biopsy, a minimally invasive procedure, allows morphometric quantification of intra-epidermal nerve fibers (IENF) most commonly expressed as the number of IENF per length of section (IENF/mm) [72,73] (Figure 63.8(a,b)). Intra- and inter-observer variability for the assessment of IENF density demonstrates good agreement [73,74] and IENFD declines with age and does not appear to be influenced by weight or height [75]. An international consortium of investigators has recently compiled a normative database for IENFD in 550 healthy participants and shown an effect of age, but no influence of height, weight or BMI [76]. The blister technique is an alternative less invasive procedure which assesses innervation of the epidermis alone and shows good agreement with punch biopsy [77].

**Diagnostic yield of IENF quantification**

No study assessing the sensitivity and specificity of IENF in DSPN is available. However, several studies in SFN have included patients with DSPN. In 58 patients with pure SFN, a cut-off IENF density of ≤8.8 per mm at the ankle was associated with a sensitivity of 77.2% and a specificity of 79.6% [78]. Similarly, in 67 patients with pure SFN a sensitivity of 88.0% and a specificity of 88.8% have been reported [79]. In a study of 210 patients with SFN, which included 65 diabetic patients, the Z-scores and 5th percentile provided the highest specificity (98% and 95%, respectively) but a very low sensitivity (31% and 35%, respectively) compared to the ROC analysis (specificity 64%, sensitivity 78%) [80]. These findings suggest that the diagnostic yield of skin biopsy may depend on the reference and cut-off values selected and the definition of SFN adopted. IENF density correlates inversely with thermal thresholds. Some have reported a closer correlation with warm and heat-pain
thresholds [78,81–83] compared to cooling thresholds [84,85] whilst others have reported the opposite, with a closer correlation with cold rather than heat detection thresholds [86,87]. A recent study has demonstrated no correlation between IENFD and the neuropathy symptom score, but interestingly an inverse correlation was demonstrated with the severity of pain assessed using the VASmax [88]. The correlation between quantitative sensory testing and IENF density therefore remains controversial.

The American Academy of Neurology, American Association of Neuromuscular & Electrodiagnostic Medicine, and American Academy of Physical Medicine and Rehabilitation have concluded that skin biopsy may be considered for the diagnosis of DSPN, particularly SFN, with a level C recommendation [89]. More recently, under the auspices of the European Federation of Neurological Societies and the Peripheral Nerve Society, revised guidelines on the use of skin biopsy concluded that IENF density is a reliable and efficient technique to confirm the clinical diagnosis of SFN with level A recommendation [90].

Additional morphologic features of IENFs include the branch density, length and mean dendritic length; all show an early reduction which progresses with increasing neuropathic severity [11,91]. Several studies with serial skin biopsies in patients with SFN have shown that axonal swellings predict a decline in IENF density [92–94]. However, they occur not only in patients with SFN [95] but also in normal individuals [96] and isolated swellings with normal IENF densities have been observed in a variety of other neuropathies [96–99].

**Diabetic neuropathy**

In patients with diabetic neuropathy, the prevalence of abnormal NC, QST, and IENF was comparable [88]. However, IENF density was significantly reduced in patients with normal NC, suggesting early damage to small nerve fibers [10,12]. Although, a recent study has shown comparable abnormalities in electrophysiology thermal thresholds and loss of IENF in diabetic patients with mild neuropathy [88]. There is an inverse correlation between IENF density and the severity of DSPN, defined by the Neurological Disability Score [11,82,100] and the Neuropathy Impairment Score [12]. Additionally, IENF density appears to be lower in diabetic patients with painful compared to painless neuropathy [11,82,101]. A 1-year diet and exercise intervention program in patients with SFN and impaired glucose tolerance (IGT) led to increased IENF density [102]. However, no change was observed in 18 diabetic patients after pancreas/kidney transplantation (SPK) [103]. This may reflect the marked IENF loss at baseline [104], particularly in diabetic patients undergoing SPK and the slower regeneration rate of IENF in diabetic patients [105]. These data suggest that IENF loss is an early feature of diabetes, progresses with increasing neuropathic severity and may improve with appropriate intervention.

**Sudomotor innervation**

Recently, a novel stereologic technique has been applied in skin biopsies and showed a correlation between sweat gland nerve fiber density, neuropathic symptoms, neurologic deficits, and sweat production [106]. The same group have now developed the technique of double immuno-staining to demonstrate co-localization of sympathetic adrenergic and sympathetic cholinergic nerve fibers in cutaneous sweat glands and vasomotor and pilomotor systems [107]. An automated quantification technique has also been developed to expedite the quantification of sudomotor innervation [4].

**Corneal confocal microscopy**

Corneal confocal microscopy (CCM) is a noninvasive ophthalmic technique that enables the *in vivo* visualization of
the sub-basal nerve plexus in Bowman’s layer of the cornea and has been shown to detect small sensory corneal nerve fiber loss in diabetic neuropathy (Figure 63.9(a,b)) [108], idiopathic small fiber neuropathy [109], Fabry disease [110], and CMT1A [111]. Corneal nerve fiber damage correlates with IENF loss and severity of neuropathy in diabetic patients [11,112] and recently three independent tests of small fiber function [113], and appears more marked in patients with painful diabetic neuropathy [11]. A correlation between loss of corneal nerve fibers and the stage of diabetic retinopathy has also been demonstrated [114]. CCM may also be more sensitive than IENFD in detecting early damage [11] and repair after simultaneous pancreas–kidney (SPK) transplantation [104,115]. Thus corneal nerve fiber density improves 6 months after combined pancreas/kidney transplantation [115] and has more recently been shown to improve before neurophysiology, QST and IENFD, providing an early surrogate marker for repair in clinical trials of human diabetic neuropathy [116]. CCM has been shown to have high reproducibility [117,118], with reasonable sensitivity and specificity [119]. To enhance the ability and hence speed of quantification of small fiber pathology, an automated image analysis system has also been developed recently to rapidly quantify corneal nerve pathology [120,121]. Furthermore, a progressive loss of corneal sensation occurs with increasing severity of somatic neuropathy, providing a functional correlate of corneal nerve fiber loss in diabetic patients [122–124]. Therefore CCM may be an ideal technique to assess alterations in small nerve fiber pathology in relation to PDN and progression or regression of neuropathic deficits [125]. These properties provide a compelling argument for CCM to be established as an FDA surrogate endpoint in clinical trials of human diabetic neuropathy.

References


Autonomic neuropathy

Chinmay S. Marathe, Liza Phillips and Michael Horowitz
Royal Adelaide Hospital/Centre of Research Excellence (CRE) in Translating Nutritional Science to Good Health, Discipline of Medicine, University of Adelaide, Adelaide, SA, Australia

Key points

- Autonomic neuropathy in diabetes is common but generally under-recognized, its manifestations diverse, and pathophysiology incompletely understood.
- Simple, inexpensive tests of cardiac autonomic function should be used widely.
- Cardiac autonomic neuropathy represents a major risk factor for mortality.
- Gastrointestinal autonomic neuropathy not only affects quality of life as a result of gastrointestinal symptoms, but impacts on glycemic control.
- Current management of the manifestations of autonomic neuropathy remains suboptimal.

Introduction

Diabetic autonomic neuropathy (DAN) may be defined as dysfunction of the autonomic nervous system in association with diabetes, or prediabetes, following the exclusion of other causes [1,2]. The autonomic nervous system innervates every organ in the body (Figure 64.1) and DAN may, accordingly, manifest as a diverse range of symptoms and deficits, including tachycardia, orthostatic and/or postprandial hypotension, exercise intolerance, nausea, vomiting, constipation, diarrhea, urinary retention and/or incontinence, sexual dysfunction, sweating abnormalities, increased glycemic variability, visual disturbance, and impaired awareness of hypoglycemia. There is a lack of standardization of diagnostic algorithms for the manifestations of DAN and, in general, screening for DAN has not been routine in clinical practice, which it should be. Despite the substantial adverse effects of DAN on both morbidity and mortality, it is arguably the least understood complication of diabetes, and current management options are limited [3]. This chapter focuses on the relevance of DAN, particularly in relation to the cardiovascular, gastrointestinal, and urogenital systems and the response to hypoglycemia; sudomotor function is addressed briefly. Broad considerations relating to the epidemiology and pathophysiology of DAN are provided initially, followed by a discussion of each of the major systems affected by DAN.

Epidemiology of Diabetic Autonomic Neuropathy

The prevalence of DAN remains uncertain given that it has been reported to range from 7.7% to 90% of the diabetic population [1,4]; this wide variation reflects inconsistencies in the definition of DAN, diagnostic methods (including their standardization) and the populations studied [1,5]. A conservative estimate would be that 20–40% of patients with diabetes develop dysfunction of the autonomic nervous system [2].

Pathogenesis of Diabetic Autonomic Neuropathy

While glycemic and metabolic control play important roles in the pathogenesis of DAN, the latter is heterogeneous and poorly defined, which has implications for effective management (Figure 64.2). So-called “autonomic imbalance” is thought to arise from a loss of parasympathetic tone and/or a transient period of increased sympathetic tone, in the setting of decreased glucose tolerance, perhaps specifically glycemic variability [4]. There is a subsequent reduction in sympathetic tone. Obesity, inflammatory cytokines, oxidative stress, dyslipidemia, and adiponectin deficiency may contribute to autonomic imbalance. For example, obesity is associated with reduced heart rate variability, apparently independent of hyperglycemia [6], supporting the concept of “prediabetic” autonomic dysfunction.
Diabetic Autonomic Neuropathy

The Diabetes Control and Complications trial (DCCT) [7,8] in type 1 diabetes (T1DM) demonstrated a reduction in DAN in the intensive control arm by 50% over a mean follow-up period of 6.5 years, reducing the prevalence of cardiac autonomic neuropathy (CAN) from 14% to 7%. A “legacy” effect was evident at 13–14 years following closeout of the study in the Epidemiology of Diabetes Interventions and Complications (EDIC) cohort with ongoing protection against CAN in the group randomized originally to intensive treatment despite comparable glycemic control [9]. In T2DM, the impact of glycemic control is less well characterized. The Steno-2 trial [10], a multifactorial intervention approach in patients with T2DM who had microalbuminuria (i.e., lifestyle modification and pharmacologic treatment to normalize blood pressure, glucose, and lipids), did, however, show a reduction in DAN after a mean treatment period of 7–8 years. Furthermore, in the Diabetes Prevention Program (DPP), in subjects at risk of diabetes who underwent intensive lifestyle changes there was an improvement in heart rate variability and a reduction in QT interval [11].

While the relevance of insights derived from models of diabetic somatic neuropathy remains uncertain, neuronal damage in DAN must ultimately be an outcome of direct neurotoxicity,
Cardiovascular autonomic neuropathy

Cardiovascular autonomic neuropathy (CAN) [25], encompassing autonomic neuropathy involving the nerves innervating the heart and the systemic vasculature, represents a major risk factor for cardiovascular disease, as well as mortality, with a greater impact than traditional cardiovascular risk factors [26]. The vagus is the longest autonomic nerve and may be the first to demonstrate pathologic changes in CAN, resulting in a degree of parasympathetic denervation and a relative increase in sympathetic tone. Accordingly, early indicators of CAN are abnormalities in the spectral analysis of heart rate variability and impaired baroreflex sensing, which progress to resting tachycardia, impaired exercise tolerance and orthostatic hypotension. There are a number of other clinical manifestations of CAN, including abnormal diurnal and nocturnal blood pressure variation, diabetic cardiomyopathy, QT interval prolongation, intraoperative cardiovascular lability, ischemic stroke and, possibly, kidney disease [27]. Despite its significance, CAN is often overlooked during assessment of patients with diabetes [28], including those with known cardiovascular disease.

Epidemiology of CAN

The reported prevalence of diabetic CAN has varied widely, but is likely to be at least 20% [29]. A European multicenter study [30] evaluated the prevalence of CAN in the outpatient setting in both T1DM (n = 647, age 11–69 years) and T2DM (n = 524, age 16–72 years). CAN, diagnosed by strict criteria (≥3 abnormal cardiovascular autonomic tests), was evident in 16.8% of patients with T1DM, and 22.1% of patients with T2DM, while the prevalence of “borderline” CAN, defined by the presence of two abnormalities on cardiac autonomic testing, was 8.5% in T1DM and 12.2%, in T2DM [30]. There is strong evidence that CAN may occur in the prediabetic state, particularly in the setting of impaired glucose tolerance (IGT) [31].

Clinical sequelae of CAN (Figure 64.3)

Mortality

CAN is associated with an increased relative risk of death [29]. For example, meta-analysis of 15 studies indicated that when compared to those diabetic subjects without evidence of CAN, the presence of a single abnormality in cardiovascular autonomic function was associated with a relative risk for death of 1.20 (95% CI 1.02–1.41), which increased to 3.45 (95% CI 2.66–4.47) when two or more abnormalities were identified. In the Action to Control Cardiovascular Risk in Diabetes (ACCORD) study, involving more than 8000 patients with T2DM, there was a strong association between CAN and both all-cause (2.14; 95% CI 1.37–3.37) and cardiovascular (2.62; 95% CI 1.49–4.91) mortality, independent of traditional cardiovascular risk factors, the duration of known diabetes and medications [28,32]. This increased mortality appears to be primarily attributable to sudden cardiac death and there is evidence for an increased risk of cardiac arrhythmias [29,33]; the QT interval is prolonged in many diabetic subjects with CAN and is an independent predictor of mortality [33]. Abnormal epinephrine signaling and metabolism secondary to an imbalance in sympathetic and parasympathetic innervation to the myocardium predisposes to arrhythmias [29]. The imbalance in parasympathetic and sympathetic tone also disrupts the normal nocturnal parasympathetic dominance, resulting in nocturnal hypertension, which is associated with increased cardiovascular mortality [29]. While impaired heart rate variability can be evidenced relatively early in the course of diabetes, resting tachycardia is now regarded as a late manifestation of diabetes with vagal impairment [4]. In patients with parasympathetic damage only, resting tachycardia (usually 90–100 beats per min, but sometimes as high as 130 beats per min) occurs, while the heart rate is likely to be around the upper limit of the normal range in patients with both parasympathetic and sympathetic damage [4]. Resting tachycardia represents an independent risk factor for cardiovascular mortality. Finally, Vinik and others have proposed a model linking CAN with myocardial dysfunction (Figure 64.3) [1].

Cardiomyopathy

The heart rate, blood pressure and cardiac output adaptations to exercise are all attenuated in the setting of CAN, leading to a reduction in exercise tolerance [29]. Diabetic cardiomyopathy,
that is, myocardial dysfunction that is not attributable to coronary artery disease, hypertension, and/or valvular heart disease, occurs frequently. Traditionally, diastolic dysfunction has been regarded as the early manifestation of diabetic cardiomyopathy [34]; however, the application of novel echocardiographic techniques and MRI technology [29] have established that changes in the left ventricle dynamics also occur. The myofibers within the myocardium spiral at an oblique angle, which results in rotation of the apical regions of the heart relative to the base during systole, a rotation referred to as “torsion” [29]. Recent studies have demonstrated that left ventricular torsion is increased in uncomplicated T1DM [35,36]. Further studies are required to define the role of CAN in the development of these left ventricular abnormalities.

**Perioperative/intraoperative risk**
Diabetes per se is associated with increased peri- and intraoperative risk, compounded in the setting of CAN which alters the hemodynamic response to induction and tracheal intubation during general anesthesia to increase the propensity for intraoperative hypothermia and cardiovascular instability [37].

**Orthostatic hypotension**
The baroreceptor reflex is a centrally mediated sympathetic response to standing, which induces peripheral vasoconstriction and reactive tachycardia. Orthostatic hypotension, defined as a decrease of greater than 10 mmHg in diastolic and/or greater than 30 mmHg in systolic, blood pressure within 2 minutes of standing [4], is a relatively late clinical manifestation of CAN [29]. Characteristic associated symptoms include dizziness, syncope, and visual impairment.

**Postprandial hypotension**
It is not widely appreciated that postprandial hypotension (PPH) (defined as fall in systolic blood pressure of ≥20 mmHg within 2 hours of a meal) occurs frequently in T1 and T2DM, particularly those with CAN, and may be more common than orthostatic hypotension [38]. PPH is associated with a number of adverse sequelae, particularly falls and cardio/cerebrovascular disorders [38]. Postprandial blood pressure control has recently been shown to be dependent on the rate of gastric emptying, such that relatively more rapid gastric emptying results in a greater fall in systolic blood pressure in patients with T2DM [39].

**Assessment of CAN** (Tables 64.1 and 64.2)
No single test is definitive for CAN. Simple, noninvasive cardiovascular autonomic reflex tests, first described by Ewing and Clarke, are used widely for the assessment of CAN [40]. Of these, changes in the R-R interval on the ECG in response to deep breathing, the Valsalva maneuver and postural change assess parasympathetic, while blood pressure changes following orthostatic provocation and the Valsalva maneuver assess sympathetic, function. These tests are generally regarded as the “gold standard” for clinical assessment of autonomic function because of their simplicity and accuracy—the heart rate response to deep breathing may have the greatest specificity [4]. The Valsalva maneuver should be avoided in patients with proliferative retinopathy. Defined reference values must be age-matched and it has been suggested that combining these tests with assessments of sudomotor function (to be discussed) may improve the precision of diagnosis [2]. While both the staging and diagnostic criteria remain to be standardized, the Toronto Expert Panel on Diabetic Neuropathy suggested
that at least two abnormal heart rate response tests be present for a definite diagnosis of CAN [2] and that the presence of orthostatic hypotension is indicative of severe CAN [29].

There are a number of other features of CAN, which are not traditionally assessed in the clinical setting. For example, both non-dipping of blood pressure, measured by 24-hour ambulatory blood pressure recording [41] and prolongation of the QT interval [33] may be markers of CAN. Spectral analysis of the ECG, a relatively novel method of evaluating CAN, transforms sequential R-R intervals into a sinusoidal function, which is analyzed to provide information about short- and long-term oscillations in heart rate variability [29]. Such tests must be performed under standardized conditions to minimize potential confounders including age, heart rate, food consumption, caffeine, smoking, the time of day, volume status, body position, and mental stress [29]. Tests of baroreflex sensitivity assess cardiac vagal tone and the sympathetic baroreflex activity assess cardiac vagal tone and the sympathetic baroreflex sensitivity [29].

Tests of baroreflex sensitivity

<table>
<thead>
<tr>
<th>Table 64.1 Assessment of diabetic autonomic neuropathy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cardiovascular Autonomic Neuropathy</strong></td>
</tr>
<tr>
<td>Cardiovascular Autonomic Function Tests [40]</td>
</tr>
<tr>
<td>Change in R-R interval on ECG</td>
</tr>
<tr>
<td>Deep breathing</td>
</tr>
<tr>
<td>Valsalva</td>
</tr>
<tr>
<td>Postural changes</td>
</tr>
<tr>
<td>Change in blood pressure</td>
</tr>
<tr>
<td>Valsalva</td>
</tr>
<tr>
<td>Postural changes</td>
</tr>
<tr>
<td>Spectral analysis ECG</td>
</tr>
<tr>
<td>Resting heart rate</td>
</tr>
<tr>
<td>Heart rate variability</td>
</tr>
<tr>
<td>Time domain measures</td>
</tr>
<tr>
<td>Standard deviation of all normal R-R intervals</td>
</tr>
<tr>
<td>Difference between longest and shortest R-R intervals</td>
</tr>
<tr>
<td>Root mean square of the difference of successive R-R intervals</td>
</tr>
<tr>
<td>Standard deviation of 5 min average of normal R-R interval</td>
</tr>
<tr>
<td>Tests of baroreflex sensitivity</td>
</tr>
<tr>
<td>Scintigraphic imaging (sympathetic)</td>
</tr>
<tr>
<td>Muscle sympathetic nerve activity</td>
</tr>
<tr>
<td>Head-up-tilt-table testing</td>
</tr>
<tr>
<td><strong>Gastrointestinal Autonomic Neuropathy</strong></td>
</tr>
<tr>
<td>Gastroparesis</td>
</tr>
<tr>
<td>Scintigraphy (gold standard)</td>
</tr>
<tr>
<td>Ultrasound (2D and 3D)</td>
</tr>
<tr>
<td>Breath tests (13C-octanoic acid or acetate)</td>
</tr>
<tr>
<td><strong>Genitourinary Autonomic Neuropathy</strong></td>
</tr>
<tr>
<td>Erectile dysfunction</td>
</tr>
<tr>
<td>Questionnaires: International Index of Erectile Function, Sexual Encounter Profile</td>
</tr>
<tr>
<td>Sacral response, bulbocavernous reflex</td>
</tr>
<tr>
<td><strong>Sudomotor dysfunction</strong></td>
</tr>
<tr>
<td>Thermoregulatory sweat test (TST)</td>
</tr>
<tr>
<td>Quantitative sudomotor axon reflex test (QSART)</td>
</tr>
<tr>
<td>Silastic imprint method</td>
</tr>
<tr>
<td>Indicator plaster method</td>
</tr>
<tr>
<td>Quantitative direct &amp; indirect reflex test (QDIRT)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 64.2 Indications for the cardiovascular autonomic function tests (CARTs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Diagnosis and staging of CAN in patients with T2DM (at diagnosis and annually thereafter)</td>
</tr>
<tr>
<td>2. Diagnosis and staging of CAN in patients with T1DM (5 years after diagnosis and annually thereafter)</td>
</tr>
<tr>
<td>3. Stratification of cardiovascular risk: pre-op, before physical activity, indication of selective beta-blocker, and suspected silent ischemia</td>
</tr>
<tr>
<td>4. Differential diagnosis of other manifestations of DAN (regardless of DAN duration): assess whether gastroparesis, erectile dysfunction, orthostatic hypotension, dizziness, syncope, or tachycardia in diabetic persons are due to dysautonomia</td>
</tr>
<tr>
<td>5. Evaluate the progression of autonomic failure and monitor response to therapy (e.g. continuous infusion of insulin, post pancreatic/islet cell transplant, use of antioxidants)</td>
</tr>
<tr>
<td>6. Differential diagnosis of other causes of neuropathy such as autoimmune autonomic neuropathy (CIDP: celiac disease, amyotrophy) or toxic-infectious neuropathy (alcohol, Hansen, HIV) as well as in cases where the presence of autonomic neuropathy is disproportionate to the sensory-motor neuropathy</td>
</tr>
</tbody>
</table>

Cardiac autonomic neuropathy (CAN); Type 1 diabetes mellitus (T1DM); Type 2 diabetes mellitus (T2DM); chronic inflammatory demyelinating polyneuropathy (CIDP)

Source: Rolim 2013 [112].

Management of CAN

Patients with diabetes and suspected, or confirmed, CAN should undergo cardiac evaluation prior to undertaking a new exercise regimen and cardiac stress echocardiography has been suggested as the investigation of choice in this setting [29]. Patients should also be advised to use their perceived level of exhaustion to guide maximal exercise, rather than rely on maximal heart rate parameters.

The main treatment strategy for CAN includes optimizing metabolic and glycemic control of diabetes. The diagnosis of early CAN, accordingly, provides an additional rationale for intensive blood glucose control. Although a number of pharmacologic approaches directed at the prevention or management of CAN have been evaluated, including angiotensin-receptor antagonists, metformin, C-peptide, and aldose reductase inhibitors [29], there is a lack of evidence to support their clinical application. Non-pharmacologic approaches specific to the treatment of orthostatic hypotension include behavioral response. Measurement of muscle sympathetic nerve activity is technically demanding and only applicable to research studies. Scintigraphic imaging (measuring sympathetic function) has been used in research studies to assess cardiovascular autonomic function. The Toronto Consensus Panel on Diabetic Neuropathy (2010) emphasized the importance of evaluating patients for the presence of CAN, and proposed that screening for CAN should be performed at the time of diagnosis of T2DM, and 5 years after diagnosis in T1DM [42].
modifications, such as avoidance of sudden changes in posture, lower limb stockings, small and frequent meals, maintaining adequate hydration, and limiting salt restriction. Pharmacologic treatments include approaches that increase volume status (e.g. fludrocortisone, or erythropoietin in anemic patients) and those that augment sympathetic activity (e.g. midodrine) [43]. Slowing the rate of gastric emptying and carbohydrate absorption by interventions such as guar [44,45] and the α glucosidase inhibitor acarbose [46] has been shown to attenuate postprandial blood pressure decreases and may prove useful in the management of postprandial hypotension.

Gastrointestinal autonomic neuropathy

The modulation of gastrointestinal functions (motor, sensory, and secretory) is complex, involving an interplay between the autonomic and enteric nervous systems and the “pacemaker” interstitial cells of Cajal (ICC), which are located within the enteric smooth muscle [2]. The neurotransmitter, nitric oxide, plays a pivotal role at the cellular level [2]. Gastrointestinal autonomic neuropathy (GAN) may involve any organ within the digestive system, including the esophagus, stomach, gallbladder, pancreas, and small and large intestine. Both motor and sensory dysfunctions occur [2], but there is only a weak relationship between dysfunctions in different regions of the gastrointestinal tract. It has only relatively recently been appreciated that GAN has substantial implications for glycemic control in both T1 and T2DM.

Epidemiology of GAN

Most studies suggest that there is an increased prevalence of gastrointestinal symptoms among diabetics, particularly females and patients with T1DM [2]. Bytzer et al. [47] evaluated gastrointestinal symptomatology in relation to diabetic complications and glycemic control in 1101 subjects with (predominantly type 2) diabetes and reported that, when compared to those without diabetic complications, patients with at least one diabetic complication were more likely to report constipation, gastroesophageal reflux, dyspepsia, abdominal pain, and fecal incontinence. Gastrointestinal symptoms are associated with increased psychologic stress, including anxiety and depression, as well as impaired quality of life [48].

Pathogenesis of GAN

As with other manifestations of autonomic neuropathy in diabetes, the pathogenesis of GAN is multifactorial and incompletely understood. A number of etiopathologies have been proposed, including vagal parasympathetic dysfunction, enteric neuronal loss, defects in inhibitory neurotransmission including decreased expression of neuronal nitric oxide, and abnormalities of smooth muscle and the interstitial cells of Cajal [49,50]. A greater availability of human tissue, particularly derived from the NIH funded Gastroparesis Clinical Research Consortium, and the development of animal models, have recently provided important cellular insights into GAN. Loss/dysfunction of the ICC appears to be central to the pathogenesis of severe diabetic gastroparesis, and may occur as a result of increased oxidative stress leading to a reduction in heme oxygenase-1, the enzyme which gives rise to carbon monoxide [49]. A lack of growth factors, including both insulin and IGF-1, has been implicated in smooth muscle dystrophy and fibrosis, which are now recognized to occur in severe gastroparesis [49]. Abnormalities of the phosphatidylinositide 3-kinase (PI3K) signaling pathway have also been described in animal models of diabetic gastroparesis [51].

It is now recognized that gastrointestinal motor and sensory functions are affected acutely by metabolic and glycemic changes—what remains contentious is the magnitude and consistency of these effects. Diabetic ketoacidosis is frequently associated with nausea, abdominal pain, and gastric stasis—the latter may be marked. Gastric emptying is slowed in the setting of acute hyperglycemia (blood glucose ∼15 mmol L⁻¹) [52]. Furthermore, in health and uncomplicated T1DM, gastric emptying is slower when the blood glucose concentration is at the upper limit of the physiologic rate (∼8 mmol L⁻¹) when compared to euglycemia (∼4 mmol L⁻¹). Conversely, insulin-induced hypoglycemia, is associated with a significant increase in the rate of gastric emptying [53] even in patients with CAN and gastroparesis, which is likely to represent an important counterregulatory mechanism. There is uncontrolled evidence that chronic improvement in glycemic control may normalize gastric emptying [54].

Diagnosis of GAN (Table 64.1)

Direct assessment of the gastrointestinal autonomic nervous system is usually not feasible, and abnormalities are traditionally inferred from assessment of cardiovascular autonomic reflexes and/or sudomotor function [2], although evaluation of the pancreatic polypeptide response to sham feeding has been used in some studies to assess the integrity of the abdominal vagus nerve. As will be discussed, the presence of CAN correlates relatively poorly with GI symptomatology, so GAN remains essentially a diagnosis of exclusion. While the absence of abnormality on assessment of GI motility does not exclude GAN as a cause of symptomatology, it may guide pharmacologic treatment.

The gold standard technique for assessment of gastric emptying remains scintigraphy, which has the capacity to quantify emptying and intragastric distribution of solid and liquid meal components concurrently. There have been attempts to standardize the methodology [55]. Ultrasound (2D and 3D) and stable isotope breath tests using noninvasive 13C-octanoic acid or acetate are acceptable alternatives [56]. While the absorption kinetics of paracetamol have been used widely, the technique is not recommended in view of its probable low sensitivity and capacity to evaluate gastric emptying of liquid only. Measurements of gastric emptying should ideally be performed during euglycemia (blood glucose 4–10 mmol L⁻¹)—at a minimum, blood glucose concentration should be monitored.
Gastroparesis, defined as delayed gastric emptying in the absence of mechanical obstruction [57], is the most important manifestation of GAN. As well as the implications for upper gastrointestinal symptoms, the stomach is central to blood glucose homeostasis by regulating the delivery of nutrients into the small intestine and the release of the incretin hormones, GIP and GLP-1. In the US at least, the number of hospitalizations due to diabetic gastroparesis appears to be on the rise [58]. The prevalence of gastroparesis is reported to be between 28% and 65% in longstanding diabetes, although there have hitherto not been any true, population-based, studies (Figure 64.4). Gastric emptying may also be abnormally rapid [59]. Because the magnitude of the delay in emptying is often modest, the terminology “gastroparesis” could arguably be reserved for patients with markedly delayed emptying associated with significant gastrointestinal symptoms and/or impaired glycemic control. There may be a higher prevalence of gastroparesis in females and in T1DM, when compared with T2DM [60,61]. It is important to appreciate that, in health, gastric emptying shows a wide inter-individual [62], but relatively small intra-individual [63,64], variation so that nutrients enter the small intestine at an overall rate of 1–4 kcal min\(^{-1}\), primarily as a result of inhibitory feedback arising from the small intestine. This inter-individual variation is, of course, increased in diabetes, which may account for what appears to be the little change in gastric emptying in patients with diabetes over time [64]. The relationship between delayed gastric emptying and CAN is weak [65]. The clinical features of gastroparesis include early satiety, nausea, vomiting, abdominal bloating and pain, impaired drug absorption, malnutrition and poor glycemic control, leading to impaired quality of life [2]. As with CAN, the relationship of upper gastrointestinal symptoms with the rate of gastric emptying is poor. The pathogenesis of symptoms remains poorly defined with disordered gastric motor/sensory function, acute changes in glycemia, gastric dysrhythmias, autonomic/enteric neuropathy, and psychiatric comorbidity all being invoked as relevant. Kassander, who coined the descriptive term “gastroparesis diabeticorum” in 1958, suggested that “the retention of stomach contents in a diabetic obviously may cause confusion as far as food intake and utilization are concerned” [66] and there is evidence that gastroparesis represents a risk factor for hypoglycemia in the early postprandial period in insulin-treated patients [67].

**Management of gastroparesis**

Despite improvements in understanding of the epidemiology and pathogenesis of GAN, treatment options for gastroparesis remain limited and the outcome is not infrequently, unsatisfactory. This is not surprising given that the underlying pathophysiology is heterogenous and, in general, the relationship between symptoms and gastric emptying is weak. Glycemic control and nutrition should be optimized. There is uncontrolled evidence that continuous subcutaneous infusion therapy may be beneficial [68]. Nutritional advice is currently based on anecdotal evidence and includes a reduction in portion size, increased frequency of meals, avoidance of insoluble fiber, and limiting carbonated beverages [25]. There is a rationale to minimize dietary factors known to slow gastric emptying, such as fat [25] and possibly, low-carbohydrate diets. If solid food is not tolerated, liquid meals should be introduced, given that liquid gastric emptying is usually relatively preserved. It is important to ensure that hydration, electrolyte status, and micronutrient balance are appropriate. While this is best achieved orally, more severe cases may require enteral, or parenteral, nutrition. Nausea and vomiting are common and antiemetics, including phenothiazines (e.g. prochlorperazine), 5 HT3 antagonists (e.g. ondansetron, granisetron), H1 receptor agonists (e.g. diphenhydramine), antidepressants (e.g. tricyclic antidepressants and tetracyclic antidepressants such as mirtazapine) and neurokinin-receptor antagonists (aprepitant) have...
been promoted as useful based on uncontrolled studies [25]. Prokinetics, including metoclopramide, domperidone and erythromycin, which have different mechanisms of action, remain the cornerstone of gastroparesis treatment [69] although, not surprisingly, there is a poor correlation between their effects on emptying and symptoms. Tachyphylaxis is also an issue with erythromycin, a motilin agonist, limiting its long-term usage [70] and the effect of all drugs is likely to be attenuated by acute hyperglycemia [71]. Metoclopramide, which is a first-line treatment in diabetic gastroparesis and has central and peripheral antidopaminergic (D2) and cholinergic properties, should be given 15 min prior to meals and pre-bed; liquid and/or subcutaneous formulations are preferable [25]. Metoclopramide has significant adverse effects, including restlessness, hyperprolactinemia, a prolonged QT interval and extrapyramidal movement disorders. Accordingly, dosage should be minimized and drug holidays considered—that the FDA has issued a “black box” warning in relation to tardive kinesia. Domperidone, which is not available in all countries, has comparable efficacy to metoclopramide [72], but a lower risk of adverse effects and acts predominantly as a peripheral D2 antagonist. Although cisapride is effective, use has been compromised by severe cardiac adverse events (prolonged QT interval, arrhythmia and death), so that it is now essentially unavailable [73]. The gastric hormone ghrelin accelerates gastric emptying in diabetic gastroparesis [74] and recently developed ghrelin agonists also have a gastrokinetic effect, with evidence of symptomatic improvement [75]. There have been encouraging open-label reports of the effects of injection of botulinum toxin into the pylorus, underpinned by evidence of pylorospasm in symptomatic gastroparesis [14]. However, recent controlled studies have failed to demonstrate efficacy and its use cannot be recommended [76]. Similarly, the promise of high frequency (∼12 pulses per min) gastric electrical stimulation, the Enterra device, remains unfulfilled [77,78] as the outcome of a multicenter study, involving 55 patients with refractory diabetic gastroparesis, was negative [79]. Pain is a difficult symptom to control and while opiates should be avoided, tricyclic antidepressants and gamma-aminobutyric acid (GABA) analogues (gabapentin and pregabalin) may have efficacy [80]. Surgery represents a last resort and there are no controlled studies. There is uncontrolled evidence that pancreatic transplantation may be associated with improvement in both symptoms and emptying [81].

**Modulation of gastric emptying to improve glycemic control**

The recognition that gastric emptying is a major determinant of postprandial glycemic excursions has stimulated the development of dietary and pharmacologic strategies to optimize glycemic control, particularly in T2DM, by modulating gastric emptying. Perhaps paradoxically the focus has hitherto been on concurrently slowing gastric emptying and stimulating the incretin hormone, GLP-1, as by acute administration of a whey protein preload given before a main meal [82]. Gastric emptying is of major relevance to the beneficial effects of some GLP-1 agonists on glycemic control—these drugs are now used widely in management of T2DM. Short-acting GLP-1 agonists, such as exenatide bd and lixisenatide, have a sustained effect to slow gastric emptying markedly, as does the amylin analogue, pramlintide [83]. In the case of short-acting GLP-1 agonists, the reduction in postprandial glycemia is related to the magnitude of slowing of gastric emptying which is, in turn, dependent on the baseline rate of gastric emptying [84]. Accordingly, these drugs probably do not slow emptying when the latter is already delayed. In contrast, long-acting GLP-1 agonists, such as liraglutide and exenatide once a week, do not have a major effect on gastric emptying with long-term usage, probably because there is tachyphylaxis to their initial effect to slow gastric emptying with sustained receptor stimulation [83]. These considerations are of relevance to the increasing use of the combination of basal insulin with a GLP-1 agonist in the management of T2DM [83].

**Enteropathy: diarrhea, incontinence, and constipation**

The prevalence of diarrhea, fecal incontinence and constipation all appear to be increased in diabetes when compared with the general population [2]. Diarrhea may reflect altered gastrointestinal secretory function and/or abnormalities in intestinal transit (which may be increased or decreased in GAN). Small intestinal bacterial overgrowth, which is a frequent complication of diabetic enteropathy, and celiac disease, must be excluded. The former may respond to intermittent, or chronic, administration of antibiotics and loperamide may also be useful. Constipation and fecal incontinence (the latter characteristically nocturnal) may occur concurrently. Disordered regulation of the internal anal sphincter and impairments in rectal sensation and compliance are likely to be important in the etiology of fecal incontinence [2]. Management of constipation is by traditional methods, involving laxatives, although some novel pharmacologic options are available [85].

**Genitourinary autonomic neuropathy**

**Erectile / sexual dysfunction (Figure 64.5)**

Erectile dysfunction (ED), defined as the “persistent inability to achieve or maintain an erection sufficient to permit satisfactory sexual intercourse” [86], is more common and severe in diabetic men when compared with nondiabetics [87], with a reported prevalence ranging from 35% to 75% [88–91]. ED impacts adversely on quality of life, personal relationships, and self-esteem [92]. Duration of diabetes, advancing age, macrovascular and microvascular complications, comorbidities including hypertension, obesity, dyslipidemia, hormonal deficiencies, and modifiable risk factors such as smoking, excessive alcohol intake, numerous drugs, and psychological disorder have all been associated with an increased risk of ED [1,92]. Some [93,94], but not all [95] studies suggest a direct association
between chronic glycemic control and erectile dysfunction. A strong association between autonomic neuropathy and ED [96] has been established, with the concept that disordered parasympathetic activity causes impaired relaxation of the smooth muscle of the corpus cavernosum leading to erectile dysfunction [97]. Retrograde ejaculation into the bladder is also sometimes seen. Importantly, diabetic ED is a strong marker for future cardiovascular disease, perhaps particularly silent myocardial ischemia, and evaluation of ED should, accordingly, also include a thorough cardiovascular assessment [1,92]. The initial step in the management of ED is to attempt to discriminate between organic and psychogenic causes, although
the latter may be relevant even when there is an unequivocal organic component(s). The management of ED is multifaceted and summarized as an algorithm in Figure 64.5. Validated measures should ideally be used to characterize ED symptoms. ED may be relatively less responsive to treatment in patients with diabetes when compared with nondiabetic individuals. There is limited evidence to suggest that CAN is associated with female sexual dysfunction.

**Bladder dysfunction**

Bladder dysfunction occurs frequently (25–80%) in long-standing diabetes and, in the broadest sense, can usually be regarded as a manifestation of autonomic neuropathy [1]. A recent report from the DCCT/EDIC study indicates that CAN is associated with urinary incontinence and lower urinary tract symptoms in females [98]. Both autonomic (sympathetic and parasympathetic) and spinal somatic nerves are pivotal in the neurophysiology of micturition. Early damage to the afferent nerve supply may result in impaired bladder sensation, urinary retention and a higher threshold for initiating the micturition reflex. Damage to the parasympathetic efferent nerves, which mediate contraction of the detrusor muscle, results in a poor urinary stream, hesitancy and dribbling, while damage to the sympathetic efferent nerves, which innervate the internal sphincter of the bladder, may lead to overflow incontinence. Chronic bladder dysfunction is associated with an increased incidence of lower and upper urinary tract infections and renal impairment. Lower urinary tract symptoms should be assessed using a validated questionnaire and characterized by comprehensive urodynamic testing, with exclusion of other causes, such as prostatic hypertrophy in males.

**Sudomotor dysfunction (including gustatory sweating)**

The sweat glands, which are of two types: (a) eccrine and (b) apocrine, are innervated by postganglionic, unmyelinated cholinergic sympathetic C-fibers. Sudomotor dysfunction, which may occur early in the course of diabetic neuropathy [86] predisposes to dry skin, itching, and foot ulceration [99]. A number of tests have been developed to assess sudomotor function, including the thermoregulatory sweat test (TST), the quantitative sudomotor axon reflex test (QSART) which has been used most widely, and indicator plaster and silicone impressions. The sensitivity and specificity of these tests is variable and their combination, for example TST and QSART, may improve diagnostic accuracy [86]. While it is clear that assessment of sudomotor function provides additional information about autonomic function, its clinical relevance remains to be established.

**Pupillary abnormalities**

Autonomic neuropathy can affect pupillary function which may impact on vision/light sensitivity and cause an Argyll-Robertson pupil. Tests of pupillary function should be regarded as research tools [1].

**Autonomic neuropathy and the response to hypoglycemia**

Diabetic patients with established autonomic neuropathy are at particular risk of severe hypoglycemia [100]. Severe hypoglycemia is predictive of mortality in diabetes; for example, a 3.4-fold increased risk of 5-year mortality has been observed in those patients with self-reported severe hypoglycemia [101] and iatrogenic hypoglycemia remains a substantial limitation in the aggressive glycemic management of diabetes [102]. The normal counterregulatory mechanisms against hypoglycemia (including a reduction in insulin production by pancreatic β cells, an increase in glucagon by pancreatic α cells and an increase in epinephrine production from the adrenals) are impaired in T1DM, and longstanding T2DM [103]. Furthermore, hypoglycemia has the capacity to impair the autonomic and humoral responses to subsequent episodes of hypoglycemia. Recent antecedent iatrogenic (as well as exercise-induced and sleep-induced) hypoglycemia results in defective glucose counterregulation (reflecting reduced epinephrine release) and impaired hypoglycemia awareness (reflecting a reduction in sympathetic neural responses), potentially inducing a vicious cycle of recurrent hypoglycemia, or “hypoglycemia-induced autonomic failure” [104,105]. The impact of the latter can be dramatic and management focuses on the avoidance of hypoglycemia. A reduction in the epinephrine response to hypoglycemia may occur in the presence of apparently normal autonomic function tests [106–108]. While it has traditionally been considered that autonomic neuropathy does not play a role in the pathogenesis of hypoglycemia unawareness [109–111], there appears to be some overlap in their features [1]. It is likely, but not proven, that patients with CAN are particularly susceptible to the adverse sequelae of hypoglycemia, including life-threatening arrhythmias. Accordingly, in patients with severe CAN targets for glycemic control should probably be relatively higher.

**Conclusions**

The broad clinical implications of DAN reflect the complex and integral involvement of the autonomic nervous system with major organ systems. Despite its substantial impact on morbidity and mortality, DAN remains the least understood of the complications of diabetes. Autonomic dysfunction occurs very frequently, represents a major socioeconomic burden, and is generally underrecognized. CAN is of particular relevance given its strong association with cardiovascular morbidity and mortality. The diagnosis of CAN in diabetes and prediabetes needs to be standardized. Early recognition of autonomic neuropathy is imperative and feasible using simple, noninvasive tests of cardiovascular autonomic function, which should be used widely. Additional studies are required to better define the prevalence, natural history, pathogenesis and
clinical implications of autonomic neuropathy in diabetes and prediabetes with a view to the development of novel, and more effective, strategies for management.

**Conflict of interest disclosure**
Professor Michael Horowitz has participated in Advisory Boards and/or symposia for Novo Nordisk, Sanofi-Aventis, Novartis, Eli Lilly, MSD, Boehringer Ingelheim, Satiogen, and Astra Zeneca/BMS and received honoraria for this activity.

**References**

31. Wu JS, Yang YC, Lin TS, et al.: Epidemiological evidence of altered cardiac autonomic function in subjects with impaired glucose


80 Kashyap P, Farrugia G: Diabetic gastroparesis: what we have learned and had to unlearn in the past 5 years. *Gut* 2010;59(12): 1716–1726.


Chapter 64


CHAPTER 65

Connective tissue disorders in diabetes

Arlan L. Rosenbloom
University of Florida College of Medicine, Gainesville, FL, USA

Key points

- Osteopenia occurs early in type 1 diabetes and is associated with a wide variety of metabolic abnormalities related to bone metabolism.
- Bone mineral density is generally increased in type 2 diabetes.
- Fracture risk in adults with diabetes is increased, much more so in type 1 than in type 2.
- Dupuytren disease is far more common in diabetes than in the general population, with abolition of the male predominance in those without diabetes and appears to be an autoimmune T-cell mediated disorder.
- Stiff hand syndrome, diabetic hand syndrome (carpal tunnel syndrome), flexor tenosynovitis, and adhesive capsulitis of the shoulder with shoulder–hand syndrome are all associated with pain and have unmistakable diagnostic features.
- Only Dupuytren disease and limited joint mobility/Rosenbloom syndrome (LJM/RS) are painless limitations and only LJM/RS occurs exclusively in association with diabetes.
- LJM/RS is associated with increased risk for microvascular disease.
- The frequency of LJM/RS has dropped dramatically since the 1970s, presumably as a result of improved diabetes control.

Introduction

Connective tissue disorders in diabetes encompass the major long-term complications, which can be attributed to alteration in the quantity and quality of structural macromolecules of the extracellular matrix [1]. These alterations affect vascular basement membrane in the retina, kidney, heart, and skeletal muscle; muscle fiber basement membrane (e.g. with cardiomyopathy); nervous system basement membrane; and lens capsule and crystallin. This chapter reviews the abnormalities of interstitial connective tissue involving skeleton, joints, skin, and periarticular tissues that have been considered to be associated with diabetes.

Skeleton

Osteopenia

Direct comparisons among various studies of bone integrity in diabetes are difficult. Various techniques have been used to demonstrate reduced bone mass, including conventional radiographs, computerized tomography, single and dual photon absorption densitometry, radiographic measurement of cortical width (radiogrammetry), resonance frequency of the ulna, total body neutron activation analysis, X-ray microdensitometry, and since the mid 1990s, dual energy X-ray absorptiometry (DEXA). Controls in a number of reports are from previous studies or the literature, rather than concurrent and analyzed simultaneously. Finally, there are wide differences in age groups studied, as well as in diabetes types, and distinction is not always made between type 1 diabetes (T1DM) and type 2 diabetes (T2DM) in a particular study, or between insulin use and other treatment.

Type 1 diabetes

Since the early 1990s, almost all studies have used DEXA for determining bone mineral content (BMC) or bone mineral density (BMD). In comparable size and age adult populations with T1DM, BMD has been reported to be normal [2] or <2 SD from normal [3], in both studies correlating with duration and control measures. Other studies in adults with T1DM generally indicate modest reduction in BMD, either independently, or related to renal function, diabetes control, oral contraceptive use, duration of diabetes, or menopausal status [4]. In both children and adults, reduced BMD has been reported soon after diagnosis [4–6]. A specific mechanism for bone mineral loss in T1DM remains elusive. In the reported studies where diabetes control has been
differentiation of preosteoblasts to mature osteoblasts [11]. Advanced glycation end-products (AGE) on type 1 collagen, formation has been suggested by an examination of the effect of advanced glycation end-products (AGE) on type 1 collagen, the predominant matrix protein of bone. In vitro studies have indicated that glycated collagen has reduced ability to promote differentiation of preosteoblasts to mature osteoblasts [11].

Type 2 diabetes
Adults with T2DM have normal or slightly elevated BMD compared to age- and weight-matched controls, particularly at the lumbar spine, hip, and radius [4]. A contributor to increased BMD in patients with T2DM has been noted in patients with T2DM who are taking HMG-CoA reductase inhibitors [12].

Fracture risk
Increased fracture risk has been noted for both T1DM and T2DM patients. Meta-analysis of studies that included over 1 million subjects has demonstrated an odds ratio (OR) for hip fractures in adults with T1DM of 6.3–6.9, but only 1.4–1.7 for those with T2DM. For spine fracture, the OR was 2.5 with T1DM. There was a 12-fold increase in hip fracture risk with nephropathy in T1DM. In the Women’s Health Study, the OR for hip fracture was 12.3 for those with T1DM and 1.7 for patients with T2DM. Older patients with diabetes are at particular fracture risk because of greater likelihood of falling with visual impairment, neuropathy, cardiac arrhythmia, and age-related sarcopenia [4].

Hyperostosis
Increased BMD in T2DM in a number of studies may be related to the association of this disease with diffuse idiopathic skeletal hyperostosis (DISH), which may involve the spine as hyperostotic spondylosis (HS), the skull as hyperostosis frontalis interna, the pelvis as osseous condensans ilii, the pelvic or other ligaments, or as large bony spurs at the heel or elbow [13]. There may be mild stiffness on arising in the morning, but spinal mobility is preserved and symptoms are usually absent. Heel and elbow pain can occur from calcaneal and olecranon spurs in about one third of affected patients and dysphagia has been described in 16% of those with hyperostosis of the spine.

While 2–4% of the general population may have hyperostotic changes, these changes are present in ~25% of the diabetes population. Pima Indians, with the world’s highest prevalence of diabetes, also have a high prevalence of vertebral ankylosing hyperostosis, comparable to their diabetes rate of 50% [14]. Of those hyperostotic patients with spine changes, 83% are male and 30% are obese. One half of all patients with hyperostosis will have manifest diabetes or impaired glucose tolerance. Obesity and diabetes appear to operate independently in the determination of this condition. Radiography reveals preservation of the disc spaces with hyperostosis seen as a line of calcification in the right anterolateral portion of the spine, typically in the thoracic region. Radiolucency is seen between the deposited bone and the cortex of the underlying vertebral body. The principal differential diagnosis of HS is from ankylosing spondylitis, which occurs in a younger population and produces more serious problems, with morning stiffness and incapacitating loss of spinal movement. Radiographic differences are straightforward [13]. The etiology of hyperostosis in diabetes is not clear. It would be attractive to invoke growth hormone hypersecretion, in view of the similarity to acromegalic lesions, but growth hormone metabolism has been described as normal [13].

Joints

Osteoarthritis
Osteoarthritis and diabetes would be expected to be associated in T2DM through the common pathogenic factor of obesity. Additionally, there are a number of processes that have been described in vitro or with experimental diabetes, such as loss of glycosaminoglycans from articular cartilage, which may be pertinent to the pathogenesis [15]. Nonetheless, it has not been possible to establish an increased risk for osteoarthritis conferred by diabetes.

Osteolysis
Osteolysis (also known as diabetic osteopathy) of the forefoot is seen as a localized or generalized osteoporosis of the distal metatarsus and proximal phalanges. Pain is variable and there may be erythema over the joint, inevitably leading to a suspicion of cellulitis or osteomyelitis. The juxta-articular erosions may resemble those of rheumatoid arthritis and gout. The etiology of this lesion is unknown, and perfect reconstruction usually occurs spontaneously [16].

Rheumatoid arthritis
The genetic interrelationship between T1DM and rheumatoid arthritis has been studied in informative families by HLA haplotype analysis, suggesting co-inheritance [17]. Insulin resistance has been noted to be a feature of rheumatoid arthritis, together
with other features of the metabolic syndrome. The decreased insulin sensitivity has been quantitatively documented, correlating with C-reactive protein elevation and with dyslipidemia [17].

**Skin and periarticular tissue**

**Scleredema diabeticorum**
This condition occurs with a male preponderance (4:1) in middle-aged adults who have a long-standing history of poorly controlled diabetes and are obese. There is thickening of the skin with a predilection for the posterior and lateral aspects of the neck and upper back, with induration extending to the face and anterior and posterior trunk and potentially involving most of the body. Hyperplasia of collagen is seen on histologic examination, and one study has shown accumulation of glycogen in unmyelinated nerve fibers [18].

**Dupuytren disease**
Dupuytren disease (DD) refers to subcutaneous fibrosis of the palmar aponeurotic space of the hands. In those without diabetes the process is sex-linked with a 6:1 male predominance, and occurs only in people of European origin. Careful examination reveals that the true incidence of DD in diabetes approaches 40%, and, as with cardiovascular risk, the presence of diabetes abolishes the sex difference. Prevalence of DD in patients with diabetes has been reported to be from 20 to 63% compared with 13% in the general population [19]. DD is also more radial in the hands of those with diabetes than in those without diabetes, affecting the third and fourth digits predominantly instead of the fourth and fifth. Characteristic lesions may precede the development of diabetes and are present in 16% of newly diagnosed older patients [20]. DD lesions causing sufficient contracture to be removed by surgery were found to contain increased contents of water, collagen, and chondroitin sulfate, as well as increased proportions of soluble collagen and of reducible cross-links, indicating synthesis of new collagen. The lesions also showed increased amounts of type III collagen and increased hydroxylation and glycation of reducible cross-links, characteristics of granulation and scar tissues. Recent evidence places DD in the category of autoimmune T-cell-mediated disorders, which are relatively common with diabetes [21].

**Stiff hand syndrome**
This “severe and sometimes incapacitating form of vascular disease of the hands” was described by Lundbaek in 1957, in five patients with long-standing T1DM [22]. The problem typically began with complaints of tingling or burning sensations in the hand, with increasing symptoms and pain aggravated by movement, leading in two patients to invalidism. The subcutaneous tissue of the fingers and palms was stiff and hard, and in two patients there were marked nail changes. There was no muscle atrophy. Radiographic study revealed calcification of the arteries of the hand in all five cases. Skin biopsy showed very few elastic fibers, but no other changes. Only a single case report of the actual syndrome has been subsequently published, suggesting that SHS may be an exceptional aggregation of diabetes complications affecting the hand (Dupuytren disease, neuropathy, atherosclerosis, diabetic hand syndrome) rather than a distinct differential diagnostic consideration [23]. Despite the lack of further reports of this dramatic condition, the term has been erroneously applied to the markedly different limited joint mobility (LJM/RS) that occurs in young patients and is without pain or paresthesias, is nondisabling, and lacks the dramatic skin changes and radiologic findings [24].

**Carpal tunnel syndrome**
Compression of the median nerve within the carpal tunnel at the wrist is the most common entrapment neuropathy, resulting in paresthesia of the thumb, index finger, and little finger, with pain that is often worse at night. Diabetes is the most common associated disorder, accounting for 5–16% of cases [13]. The condition is frequently referred to as the diabetic hand syndrome and, as with the stiff hand syndrome of Lundbaek, the term has been erroneously applied to LJM/RS.

In persons with diabetes, carpal tunnel syndrome (CTS) may not be simply due to nerve compression but a manifestation of diabetic neuropathy with decreased conduction velocity of both median and ulnar nerves with, in addition to typical thenar muscle atrophy, atrophy of the intrinsic and hypothenar muscles. The contractures involve the metacarpophalangeal (MCP) and proximal interphalangeal (PIP) joints of all fingers equally.

A literature review of 64 articles on CTS confirmed diabetes or a family history of diabetes as a primary risk factor, along with menopausal age, obesity, lack of fitness, smoking, alcohol intake, and osteoarthritis of the carpal–metacarpal joint of the thumb [25].

**Flexor tenosynovitis**
Flexor tenosynovitis (FTS), also known as “trigger finger” or stenosing tenovaginitis, can be congenital or hereditary in children, almost exclusively involving the thumb. In adults, it is estimated that one third of multiple palmar FTS is due to diabetes, although the prevalence in those with diabetes is not known [13]. There is marked female predominance, predilection for the right hand, and preferential involvement of the thumb, middle, and ring fingers. Fibrous tissue proliferates in the tendon sheath, particularly where the tendon is constricted in its passage through a fibrous ring or pulley or over a bony prominence, with swelling distal to this constriction and pain with movement of the enlarged segment through the narrowed ring. Palpable or audible crepitus may be present with movement; locking in flexion or extension occurs with impaction of the nodule proximal or distal to the thickened segment of tendon sheath [16].

**Adhesive capsulitis of the shoulder and shoulder–hand syndrome**
Thickening of the joint capsule and its adherence to the head of the humerus results in marked reduction in the volume of the glenohumeral joint. In some cases this periarthritis may precede,
accompany, or follow diffuse swelling, coldness, erythema, tenderness, and hyperhidrosis of the hand. Subsequently, swelling and vasomotor instability resolve, with the development of trophic skin changes and contractures. After weeks or months, the tenderness, swelling, and vasomotor dysfunction completely resolve, with residual atrophic or dystrophic changes, finger contractures and, occasionally, frozen shoulder with atrophy of the shoulder girdle muscles. Osteoporosis of the bones of the hand and shoulder is seen. In addition to shoulder–hand syndrome, this has been referred to as reflex sympathetic dystrophy, causalgia, post-traumatic osteoporosis, and Sudeck atrophy [13].

Adhesive capsulitis occurs at a younger age in those with diabetes, is typically less painful, but also less responsive to treatment and longer lasting than in those without diabetes. The prevalence is estimated to be 11–30% in those with diabetes and 2–10% in the nondiabetic population [19].

Limited joint mobility (LJM)/Rosenbloom syndrome (RS)
The initial observation of striking limitation of extension and flexion of the interphalangeal, MCP, and wrist joints, in association with short stature, thick, tight, waxy skin, delayed sexual maturation, and early microvascular complications, in three older teenagers with long-standing diabetes [26], was followed by the description of milder manifestations in 28% of 229 campers with diabetes aged 7–18 years [27]. Subsequent studies have reported prevalences of 9–63% among T1DM patients, depending on the age of the population and the duration of diabetes, as well as the examination techniques. A comparison of the findings in 1998 to those done by the same examiners using the same techniques 20 years earlier found a greater than fourfold reduction in frequency of LJM/RS together with a decrease in the proportion having more advanced changes, attributed to improved blood glucose control measures [28]. However, a subsequent report from England indicated a frequency of LJM/RS in the 1990s comparable to that reported in the original description in the 1970s in the US [29]. Several reports have described LJM/RS from a broad geographical range in T2DM also, with frequencies ranging from 25 to 76% [30].

There is no influence of sex or race on prevalence. In control pediatric populations, stiffness of the little fingers only was usually found in fewer than 4%, whereas older populations without diabetes may have as much as 25% involvement, probably related to occupation and aging. A significant independent contribution of cigarette smoking to both LJM/RS and DD has been described in adults with T1DM and T2DM.

Changes begin in the MCP and PIP joints of the little finger and extend medially; the distal interphalangeal joint may also be involved. Changes are also seen in the metatarsophalangeal joints [31] and larger joints, most commonly wrist and elbow but also ankles and cervical and thoracolumbar spine [32]. The limitation is painless, and nondisabling. Involvement of the foot can contribute to abnormal foot pressures and damage [33].

Examination and classification
The original method for demonstrating milder LJM/RS was to have the patient place the hands on a flat surface palm down with the fingers fanned; the examiner would then determine contact with the plane surface by viewing at table level. Normally, the entire palmar surface of the fingers makes contact [32]. A simpler method is to have the patient attempt to approximate the palmar surfaces of the interphalangeal joints with the fingers fanned (Figure 65.1). Whether either of these approximations is possible or not, the examiner extends the proximal and distal interphalangeal and MCP joints. In addition to finding these limited to less than 180° and 60°, respectively, the examiner may find resistance to the permitted movement [31,33]. Thickening of the tissues surrounding the limited joints and inability to tent the skin may also be noted, particularly over the dorsa of the fingers and hand. The examiner also attempts to extend the wrist maximally to at least 70° and the elbow to at least 180°. The ankle should flex maximally to at least 100°, the cervical spine lateral flexion should permit ear-to-shoulder juxtaposition, and thoracolumbar spine lateral flexion should be at least 35° in young people [34].

Figure 65.1 Inability to approximate the palmar surfaces of the fingers in a 14-year-old girl with diabetes for 7 years.
This examination will reveal obvious LJM/RS but, as noted, the clinical sensation of joint limitation without absolute reduction in maximum extension is quite subjective. Buithieu et al. [35] compared MCP and wrist maximal extension in 239 controls and 211 T1DM patients aged 9–30 years using a goniometer, and found a significant proportion of those without clinically apparent limitation to actually have MCP and wrist limitation, defined as less than 2 SD of the control mean. The 172 subjects without apparent LJM/RS included 25% with MCP limitation and 13% with wrist limitation. Similar findings have been reported from France [36]. Thus, earlier detection of LJM/RS may be possible through objective measurement. This is also likely for the associated changes in the lower extremity.

Mild limitation indicates involvement of one or two PIP joints, one large joint, or only the MCP joints bilaterally. Moderate limitation refers to involvement of three or more PIP joints or one finger joint and one large joint bilaterally. Severe limitation refers to obvious hand deformity at rest or associated cervical spine involvement [34].

Because LJM/RS may involve more than hand joints, the term “cheiroarthropathy” is not adequate, leading to such incongruous statements as, “Cheiroarthropathy … may also affect the large joints” [37]. Referring to this condition as “the scleroderma-like syndrome” [38] is also not appropriate because only the more severely affected individuals will have “digital sclerosis” [32]. It has also been called “Rosenbloom syndrome” which might be useful to be sure there is no confusion with the other finger joint limitation conditions associated with diabetes, as has been frequent in the literature [39–42].

Natural history

Cross-sectional studies demonstrate that duration of diabetes is the most important variable in the appearance of LJM/RS. However, when the time of development of joint changes is determined, age attained is much more important than duration of diabetes in the development of LJM/RS. The interval between detection of mild limitation and progression to moderate or severe changes varies from 3 months to 4 years, with a mean of 2 years. After this period, progression (if any) is very slow. Most do not progress beyond mild changes [43].

There was no correlation between loss of BMC by photon absorptiometry and the presence or absence of LJM/RS in children and adolescents, which is to be expected in view of the different timing of these findings in the course of T1DM [9]. One half of young patients affected with more than 5 years’ duration of diabetes had moderate or severe limitation, defined as at least three interphalangeal joints or one finger joint and one large joint bilaterally. Approximately one third of the affected group had severe limitation, that is, involvement of all the fingers and often of the cervical spine as well [43]. In adult populations, about two thirds of patients affected had at least two fingers involved [44].

Hepatomegaly and proteinuria in many of the initial patients who were severely affected by LJM/RS and had associated thick, tight, waxy skin, suggested that LJM/RS was the result of poor control [26,32]. The only controlled study of metabolic regulation from the onset of diabetes until the occurrence of LJM/RS, in order to avoid the numerous biases of cross-sectional study, described a strong association of long-term control and the appearance of LJM/RS. For every 1% increase in average HbA1c from onset, there was an approximately 46% increase in the risk of LJM/RS [45]. As noted earlier, the greater than fourfold decrease in prevalence of LJM/RS and in severity in those few affected between 1976–78 and 1998 in the US can only be explained by the improvement in control measures [28]. A similar but less dramatic reduction in prevalence has been recorded for a young adult UK population (age 27 ± 1 years, n = 204) between the 1980s and 2002, from 43% to 23% [46].

Effects on growth

In individuals with more than 3 years’ duration of diabetes, mostly with onset before adolescence (thus permitting sufficient time for growth failure to occur), 68% of subjects without LJM/RS, were below the 50th percentile for height rather than the expected 50%, and most of this discrepancy was below the 25th percentile (38% vs. expected 25%). Mild LJM/RS resulted in four times the excess below the 25th percentile (72%), and moderate to severe limitation was not associated with significantly greater deleterious effect on stature (77% below the 25th percentile). The small proportion above the 75th percentile was similar for those without LJM (12%), and with mild or more severe LJM/RS (both 10%) [43]. As with the marked reduction in frequency of LJM/RS, the associated impairment in growth and the modest growth impairment in those without LJM/RS virtually disappeared in the ~20 years following the initial study; of 175 children with greater than 2 years’ duration before puberty and overall duration greater than 3 years, corresponding to the criteria of the earlier study, only 22% in 1998 versus 37% in 1982 were <25th percentile for height and of those few with LJM/RS, only 33% in 1998 were <25th percentile compared to 77% in 1982 [28].

Differential diagnosis (Figure 65.2)

The diagnosis of LJM/RS and its distinction from other joint problems should be straightforward, especially in young patients [30]. Nevertheless, since the original description, numerous reports have referred to the various conditions listed in Table 65.1 as “limited joint mobility,” or have considered several of these conditions to be varying manifestations of the same process, using the terms for the distinctive diagnostic conditions interchangeably, in particular referring to LJM/RS as diabetic hand syndrome or stiff hand syndrome [47,48]. LJM/RS is easily distinguishable from DD, the only other condition involving the hand that is not associated with pain, by the absence of palmar fascial thickening and nodules, and by the finger distribution. DD is not seen in young patients; in older patients, DD and LJM/RS may be seen together [44]. All the other conditions are associated with pain and other characteristic findings, as
Finger joint limitation

Pain/paresthesia?

Yes

All finger joints?

Yes

Shoulder?

Yes

Flexor tenosynovitis

(1st, 3rd, 4th fingers)

No

Foot and other joints affected?

No

5th finger affected?

Yes

LJM/RS* (Limited joint mobility/Rosenbloom syndrome)

No

Dupuytren disease

Hand muscle atrophy?

Yes

Carpal tunnel syndrome

(diabetic hand syndrome)

No

Stiff hand syndrome

*Limited joint mobility/Rosenbloom syndrome

Figure 65.2 Differential diagnosis algorithm for finger joint limitation in diabetes.

Table 65.1 Differential diagnosis of joint syndromes involving the hand in diabetes

<table>
<thead>
<tr>
<th>Condition</th>
<th>Usual fingers involved</th>
<th>Pain/paresthesia</th>
<th>Muscle</th>
<th>Frequency</th>
<th>Other distinctive features</th>
</tr>
</thead>
<tbody>
<tr>
<td>LJM/RS</td>
<td>Begins 5th, extends radially</td>
<td>No</td>
<td>Normal</td>
<td>10–20 % of T1DM after 5 years and puberty</td>
<td>Not disabling; associated with thick, tight, waxy skin</td>
</tr>
<tr>
<td>Dupuytren disease</td>
<td>3rd, 4th</td>
<td>Normal</td>
<td>1/3 associated with diabetes</td>
<td>Up to 40% over age 40; 16% at diagnosis</td>
<td>Milder expression common in women; may precede diabetes</td>
</tr>
<tr>
<td>Flexor tenosynovitis</td>
<td>1st, 3rd, 4th</td>
<td>No</td>
<td>Normal</td>
<td>5–16% due to diabetes</td>
<td>Marked female predominance; locking, disability</td>
</tr>
<tr>
<td>Carpal tunnel syndrome (diabetic hand)</td>
<td>All</td>
<td>Yes</td>
<td>Intrinsic, palmar atrophy</td>
<td>Rare: T1DM &gt;20 years</td>
<td>Ulnar as well as median nerve involvement in diabetes</td>
</tr>
<tr>
<td>Stiff hand</td>
<td>All</td>
<td>Yes</td>
<td>Normal</td>
<td>Shoulder limitation in 10–20% of older T1DM, T2DM; reflex dystrophy unusual</td>
<td>Disability; calcified vessels; hard palmar skin, soft dorsal</td>
</tr>
<tr>
<td>Shoulder–hand syndrome (reflex dystrophy)</td>
<td>All</td>
<td>No</td>
<td>Atrophy</td>
<td>Most bilateral; often associated with other hand syndromes</td>
<td></td>
</tr>
</tbody>
</table>

noted in Table 65.1. Rheumatoid arthritis is rarely associated with diabetes, and patients with LJM/RS do not meet the pain or inflammatory criteria for rheumatoid arthritis [17,30].

The superimposition of other hand syndromes on LJM/RS as patients age is not surprising, especially with prolonged duration. The presence of pain or paresthesia, neurologic findings, disability, finger locking, swelling, muscle atrophy, palmar skin or fascia thickening, or absence of greater involvement of the ring and little fingers could be indicators of additional problems. One would expect people with LJM/RS to be at
greater risk for other connective tissue proliferative problems and neuropathy. As noted earlier, generalized scleroderma can rarely be associated with diabetes, but should provide no diagnostic confusion. In older patients, osteoarthritis and occupation-related limitation should be easily distinguished by history and physical examination.

**Neuropathy**

Kennedy et al. [49] noted prolonged nerve conduction velocity for the ulnar and median nerves in older patients with LJM/RS, compared with those without limitation and comparable durations of diabetes. Also noted was decreased vibratory sense in both upper and lower extremities in the presence of LJM/RS. Starkman et al. [50] found that LJM/RS was associated with a 4.3-fold relative risk of clinical neuropathy in T1DM.

**Pulmonary changes**

The association of limited pulmonary capacity with LJM/RS was first noted by Barta in a single patient [39]. A study of 12 adolescents and young adults with severe LJM who were matched for age, sex, height, and duration of T1DM to 11 without LJM/RS, found that LJM/RS subjects had significantly less total lung capacity, thoracic gas volume, residual volume, forced vital capacity, and forced expiratory volume [51]. Similar findings have been reported from Hungary and Cuba. What cannot be ascertained is whether these individuals had limited pulmonary compliance, decreased mobility of the chest wall, or both.

**Skin changes**

Thick, tight, waxy skin, most prominent over the dorsum of the hands and the forearms, was noted in the initial patients and in about one third of those with LJM/RS described subsequently. This change was apparent only in those with moderate to severe LJM/RS [30,32]. In a study of 375 patients aged 7–25 years, skin changes were present in 34% of those without LJM/RS, in 70% of those with mild LJM/RS (involving only the fifth PIP joint) and in 100% of those with more severe limitation [52].

Ultrasound studies of 92 T1DM patients aged 20–38 years, with a wide range of disease durations, demonstrated that those with LJM/RS had thicker skin than those without limitation, who in turn had thicker skin than those without diabetes [53]. Skin thickness assessed by ultrasound measurement was also noted to correlate with LJM/RS in 80 children and adolescents [54].

Biopsy studies have shown thickening of the dermis and epidermis with accumulation of collagen and loss of skin appendages compared with control biopsies [30,50]. The pathologic features have been more carefully defined by Hanna et al. [55]. They found clinical evidence of skin thickening in 22% of T1DM patients and 4% of controls. Full-thickness skin biopsy specimens from the forearm were analyzed in nine patients with T1DM and thick skin, four patients with T1DM and clinically normal skin, four patients with progressive systemic sclerosis, and four normal control subjects. They defined distinct differences from sclerodermatous and normal individuals: the T1DM patients with thick skin showed active fibroblasts and extensive collagen polymerization in the rough endoplasmic reticulum. Unlike scleroderma, in which there is bimodality of collagen fiber sizes, the thick skin of diabetes demonstrated predominance of large fibers, a finding that was also present in those with diabetes who did not have thick skin. Thus, the fibrosis in these two conditions appears basically different. Biochemical studies are discussed in the following sections.

**Association with microvascular disease**

The initial seven patients with severe LJM/RS involving fingers and large joints included five with clinically apparent retinopathy or proteinuria before the age of 18 years [27,32]. In a unique longitudinal study of a clinic cohort in which the onset of LJM/RS and of microvascular complications could be documented, who had diabetes duration of more than 4.5 years (the shortest duration at which microvascular complications were noted), 82 of 169 patients had LJM/RS, of whom 41 also had microvascular complications, in contrast to only 10 of the 87 patients without LJM/RS. Severity of the joint limitation correlated directly with the frequency and severity of the microvascular disease. Actuarial analysis indicated an 83% risk for microvascular complications after 16 years of diabetes if joint limitation was present, but only a 25% risk in the absence of LJM/RS. The differences between the LJM/RS and non-LJM/RS groups could not be accounted for by differences in patients’ ages or duration of diabetes. The 4.5-fold greater likelihood of microvascular disease in the group with joint limitation in this population indicated that LJM/RS identified a group exceptionally at risk for the development of early complications [34].

Among adult Irish patients with T1DM, retinopathy was present in 52.4% with LJM/RS compared with only 12.3% in those without joint limitation. Those with LJM/RS and retinopathy had much longer durations of diabetes than those with LJM/RS and no retinopathy. To assess whether the presence of LJM/RS was a risk factor for the development of retinopathy, independent of duration of diabetes, two subgroups of 20 patients each, with and without LJM/RS, were matched for duration. Retinopathy was much more common in the group with LJM/RS (85% vs. 40%), with a highly significant difference in the prevalence of proliferative retinopathy—70% in patients with LJM compared with 15% in patients with normal joint mobility [56].

A similar attempt was made to correct for duration of diabetes in a young population of 311 subjects aged 6–27 years undergoing fluorescein angiography to determine the prevalence of early retinopathy. The relationship between LJM/RS and retinopathy was highly significant (p < 0.0001), and this was not due to the effect of duration on both complications: the interaction of LJM/RS, retinopathy, and duration was insignificant. The presence of LJM/RS was predictive of associated retinopathy at the level of clinical recognition (more than 10 microaneurysms);
43% of those with LJM/RS and more than 4 years’ disease duration had retinopathy compared with 15% of a matched duration group without LJM/RS [57]. In the 45% of 110 Ethiopian patients with T1DM who had LJM/RS, retinopathy was found to be twice as frequent as in those without LJM [58]. Among 150 Joslin Clinic patients under 40 years old, of whom 75 had LJM/RS, the relative risk of retinopathy conferred by the presence of joint limitation was 3.7, and of preproliferative and proliferative retinopathy, 4.2 [50]. The largest survey reported, involving 375 persons aged 7–25 years, noted positive correlations between skin involvement, severity of joint limitation, and diabetic retinopathy, although the latter was evaluated only through the undilated pupil [50]. Another large study, of 357 subjects, noted significant association of LJM/RS with retinopathy and proteinuria [59]. Other studies of T1DM come from Spain, of patients aged 18–57 years, in which correction was made for age and duration of diabetes, in which LJM/RS was significantly associated with albuminuria and retinopathy and of 335 patients from Germany aged 14–40 years, 34% of whom had LJM/RS, which was significantly associated with higher albumin excretion rates and hypertension. LJM/RS was also an independent predictor of retinopathy only in men, and of macrovascular disease (increased intima-media thickness of the carotids and atherosclerotic plaques) only in women [30]. In a study of 12 adolescent patients with proteinuria or hypertension and LJM/RS, the increase in the mesangial matrix and basement membrane thickening on renal biopsy correlated with the severity of joint limitation [60].

Significant correlations between LJM/RS and microvascular disease have been reported with T2DM, but these correlations have been less impressive than in T1DM. This difference is particularly apparent in studies that involve both T1DM and T2DM [44,58].

**Biochemical studies**

Increased accumulation of dermal collagen that is relatively insoluble and resistant to enzymatic digestion has been recognized as a characteristic of connective tissue aging in both T1DM and T2DM. Skin biopsies from the lateral thigh of 23 patients with T1DM yielded 8 with less than 2.5% collagen.

In a study of 12 adolescent patients with proteinuria or hypertension and LJM/RS, the increase in the mesangial matrix and basement membrane thickening on renal biopsy correlated with the severity of joint limitation [60].

**Constitutional versus metabolic considerations**

The intuitive attribution of LJM/RS and associated skin changes to metabolic control along with the other long-term complications of diabetes has received extensive experimental support since publication of the DCCT results. The definitive demonstration of the importance of long-term diabetes control in the emergence of LJM/RS [45] and the marked decrease in prevalence of this complication with improved control measures since the 1970s [28] have indicated that the findings of the DCCT can be extended to the specific complication of LJM/RS.

Although no differences in HLA distribution between those with and without LJM/RS have been described [43,49], and there is no difference in frequency of organ-specific autoimmunity [43], a genetic component has been suggested by the observation of joint limitation in nondiabetic relatives of T1DM patients. Brice et al. [68] found 42% of 112 T1DM patients aged 2–12 years to have LJM, with 78% of those with duration of diabetes of more than 7 years being affected; this
is three to four times the prevalence reported by others for this largely preadolescent age group. This diagnostic exuberance may explain their finding among 214 nondiabetic first-degree relatives, that 35% of the relatives of children with LJM/RS were similarly affected as were 13% of relatives of those children without LJM/RS. Rosenbloom and colleagues [69] examined 204 T1DM patients aged 7–23 years and 336 of their first-degree relatives. They also examined simplex and multiplex pedigrees with T1DM and normal controls. Only 1 of the non-diabetes-related normal controls had fifth finger joint limitation; 3% of 225 nondiabetic parents were affected compared with 21% of the probands. Of the 108 nondiabetic siblings, only 1 had fifth finger joint limitation. Three parents had adult-onset diabetes and had joint limitation. Of the nondiabetic relatives with joint limitation, none was related to a proband with LJM/RS and all tested were negative for islet cell antibodies. Among 11 T1DM multiplex families with at least 1 member having joint limitation, the concordance rate for LJM/RS was no greater than expected for age and duration of diabetes. Thus, the evidence for LJM/RS being a metabolic consequence of diabetes included the absence of limitation among nondiabetic first-degree relatives of probands, including of probands with LJM/RS, and that concordance for joint involvement was not increased in first-degree relatives with T1DM.

Further evidence for a metabolic explanation for LJM/RS in T1DM comes from its description in non-HLA-associated, nonautoimmune T1DM, in two siblings with the syndrome of diabetes insipidus, diabetes mellitus, and optic atrophy [70], as well as in two siblings with diabetes from infancy due to pancreatic hypoplasia [71]. Tissue culture studies provide the opportunity to examine genetically determined donor characteristics generations of cells removed from the donor metabolic milieu. The viability of cultured fibroblasts from two patients with severe LJM/RS and growth failure was found to be completely normal for age, indicating that there is not an inherent cellular defect in growth in these youngsters [72].

The essential (apparently) metabolic derangement for the development of LJM/RS and associated skin findings as well as other complications of diabetes that occur despite the type of diabetes (T1DM, T2DM, secondary) must require a constitutional predisposition, which is not specific. Monnier and colleagues [73] proposed a means by which this constitutional factor may be expressed. They noted that the age-related collagen-linked fluorescence in skin biopsies from long-standing T1DM reflecting AGEs was, as previously noted, greatly increased compared with controls. The rate of browning was not significantly different from normal in those without retinopathy but was 2.5 times greater in the presence of retinopathy. These findings suggested that there was a mechanism controlling the browning rate of collagen in those individuals who did not develop retinopathy.

An alternative metabolic pathway for ketoamine products could provide such a mechanism [74]. The oxidation of earlier glycation products to form relatively inert carboxymethyllysine rather than AGEs would reduce the ill effects of protein glycation. Differences in response to this process could explain different susceptibility to the development of complications.

Conclusion

The abnormalities in bone, joints, skin, and periartricular tissue described in this chapter cannot be attributed to a single pathogenic mechanism. As osteopenia is seen early in the course of diabetes and does not appear to worsen with duration, a basic metabolic abnormality concomitant with the development of hyperglycemia and unlikely to be related to other complications is probable. Vascular insufficiency associated with vessel calcification can explain the stiff hand syndrome, and neuropathy the carpal tunnel syndrome as it appears in diabetes, where it is referred to as the diabetic hand syndrome. The lesions attributable to connective tissue proliferation, involving palmar fascia (DD), the periartricular tissue (adhesive capsulitis of the shoulder, LJM/RS), the tendon sheaths (flexor tenosynovitis), and the skin, may have the most relevance to the long-term disabling or fatal complications of diabetes in which connective tissue of basement membranes is abnormal.

Although the lesions of diabetes complications are frequently compared to those of aging, the skin and periartricular lesions (and perhaps the osteopenia) of diabetes either are unique or differ substantially from comparable lesions in nondiabetic individuals. This is also true for the complications of the retina, the heart, the kidneys, and the peripheral nerves. LJM/RS is of particular interest because of its early onset, the associated skin and lung changes, the established relationship to microvascular disease, and the dramatic reduction in its frequency in T1DM during an era of improved ability to control glycemia.

References


Spagnola AM, Bennett PH, Ternsald PI: Vertebral ankylosing hyperostosis (Forestier's disease) and HLA antigens in Pima Indians. *Arthritis & Rheumatology* 1978;21:467–472.


CHAPTER 66
The diabetic foot

Frank L. Bowling and Andrew J.M. Boulton
University of Manchester, Manchester Royal Infirmary, Manchester, UK

Key points
• Peripheral neuropathy is concurrent in 90% of foot ulcers and is a major contributor to diabetes-related foot ulcer development.
• Renal disease increases risk of foot ulceration and dialysis is a strong predictor of lower extremity amputation.
• A comprehensive diabetic foot examination is essential for appropriate assessment, prevention, and management of diabetic foot complications.
• Diabetes-related foot ulcers do not require cultures or antibiotic treatment in the absence of clinical signs of infection.
• Off-loading using a total contact cast is a highly effective treatment for diabetic foot ulcers and acute Charcot neuroarthropathy.

Introduction
There are currently an estimated 360 million people suffering with diabetes mellitus worldwide which equates to approximately 5% of the global population. Foot ulceration is a common complication of diabetes and at any one time 2–4% of the diabetes population is likely to have an active foot ulcer and more than half of these will become infected [1]. Eighty-five percent of infected diabetic foot ulcers progress to lower limb amputation after which 5-year mortality rates are significantly increased.

Many of the complications associated with diabetes are inextricably linked through abnormalities in microvascular and macrovascular function underpinned by the metabolic anomalies resulting from chronic hyperglycemia. As a consequence of the multisystem nature of diabetes, the development of foot lesions and subsequent healing will be influenced by the extent and severity of other complications including peripheral vascular disease, cardiovascular disease, cerebrovascular disease, retinopathy, renal failure, depression and a significantly reduced quality of life.

Management of diabetic foot ulcers (DFU) is a long and complex process that begins before ulcer formation through a program of prevention. The last decade has seen the development, dissemination and implementation of consensus guidelines which clearly illustrate the interventions necessary to tackle the issue of diabetic foot ulceration. However, the increasing prevalence of diabetes and escalating healthcare costs associated with managing the complications make it more important than ever that these guidelines are understood and implemented appropriately.

The development of a foot ulcer is not an inevitable consequence for patients with diabetes. Lesions arise from a combination of multiple factors and processes which can be modified in order to reduce risk of ulceration. This chapter will explore ulcer pathogenesis, risks for ulceration, and the assessment process as a means of reducing risk. Management of infected foot ulcers will be included in a discussion of general management of DFUs followed by Charcot neuroarthropathy (CN).

Incidence and prevalence of diabetic foot ulcers
A person with diabetes has a lifetime risk of foot ulceration of between 15 and 25% [2]. The North-West Diabetes Foot Care Study identified the annual incidence of new foot ulceration in community diabetes patients as 1.7% [3].

A DFU is the result of the interaction of multiple physiologic processes with a large number of risk factors. Peripheral vascular disease and distal symmetrical sensorimotor neuropathy are highly influential in the pathogenesis of ulcer formation. The identification of relevant risk factors is essential for all patients with diabetes.

Peripheral vascular disease
The frequency of occurrence of peripheral vascular disease (PVD) in diabetes mellitus has been established by a number
of prospective longitudinal studies. The diabetes audit and research study in Tayside Scotland (DARTS) reported the annual incidence of PVD in patients with type 1 diabetes (T1DM) was 5.5 per 1000 patients and 13.6 per 1000 patients for those with type 2 diabetes (T2DM) [4].

Diabetes was reported to be positively associated with PVD (odds ratio 2.8) in the National Health and Nutrition Examination Survey with 4.3% prevalence in the general population of the USA [5].

A community-based study of T2DM patients in Australia identified the prevalence of PVD at entry to the study to be 13.6% [6]. The incidence of new PVD at follow-up 5 years later was 3.7 per 100 patient-years. In addition PVD was identified as a strong risk factor for cardiac death and independently associated with raised total serum cholesterol and smoking.

With respect to foot ulceration in Europe, the Eurodiale study reported that peripheral arterial disease was present in approximately half of all patients with foot ulcers assessed at baseline which the authors reported to be a shift away from ischemic ulcers towards neuroischemic ulcers. Twelve percent of all patients had critical limb ischemia [7]. A study by Gershater reported that 45% of patients with a diabetes-related foot ulcer also had signs of severe neuroischemia characterized by systolic ankle pressure of 80 mmHg or a systolic toe pressure of 45 mmHg [8].

The specific role of PVD in ulcer pathogenesis is difficult to determine as patients are rarely involved in health care during the early, asymptomatic stages of the disease. The significance of this is illustrated in findings from a study investigating factors related to outcomes in neuropathic and neuroischemic foot ulcers. Ulcers diagnosed as neuroischemic following objective investigation for the study would not have been considered as neuroischemic in the absence of the objective testing [9]. However, microvascular defects such as arteriolar hyalinosis, sclerosis and basement membrane thickening are observed in diabetes subjects. Large vessel involvement (iliac artery, femoro-popliteal and infra-popliteal) as seen with occlusive disease may contribute to ulcer formation due to subsequent reduced foot perfusion.

Patients with PVD are reported to have prolonged healing times, increased rates of amputation, higher mortality, and more disabling comorbidities. Neuroischemic ulcers tend to be larger with increased deep tissue and bone involvement in comparison to neuropathic foot ulcers. The apparent differences in clinical characteristics and outcome have prompted the suggestion that neuroischemic ulcers may be an entirely different disease entity. Further epidemiologic studies are necessary to explore this possibility [7].

Foot ulcers in patients with PVD also appear to be at increased risk for infection. The presence of a foot ulcer with or without infection increases perfusion demand and the imbalance in supply may decrease the likelihood of healing.

Peripheral vascular disease clearly plays a significant role in diabetic foot ulceration the specifics of which remain unclear. However, reducing the risks of developing PVD through education, smoking cessation, increased physical exercise, and healthy diet is likely to be of long-term benefit.

**Diabetic peripheral neuropathy**

Neuropathy describes the progressive loss of sensation as a result of axonopathy and demyelination of small and large fiber nerves. C-fibers contribute to pain and temperature sensation whilst A-delta fibers carry information about vibration and joint position sensation. The mechanisms behind neuropathy are thought to originate with multiple metabolic pathways arising as a result of hyperglycemia and dyslipidemia and it has been well demonstrated that intensive diabetes treatment can prevent further deterioration in complications such as peripheral neuropathy [10].

Diabetic peripheral neuropathy (DPN) can also affect the central nervous system as well as the peripheral nervous system and studies have demonstrated neurologic changes within the brain related to diabetes [11]. Furthermore, responses to pain in patients with DPN have been explored and compared against subjects with diabetes but no DPN and nondiabetic controls. Functional magnetic resonance imaging (fMRI) illustrated that there were significantly different patterns of activation in the cerebral cortex of patients when nociceptive heat stimuli were applied to the foot and thigh [12].

There are a wide range of risk factors for developing neuropathy and cardiovascular disease plays a major contributory role. The presence of cardiovascular disease and diabetes is also associated with increased mortality. Other risk factors related to neuropathy include increasing age, duration of diabetes, and poor glycemic control.

The development of neuropathy is not a foregone conclusion given that a number of the risk factors are open to modification. The possible health benefits associated with strict glycemic control were explored in the Diabetes Control and Complications Trial Research Group (DCCT) which demonstrated significant reduction in progression of complications including neuropathy, nephropathy, retinopathy, and cardiovascular disease [13].

Diabetic neuropathy exists in many forms depending on the distribution and symmetry of the pathologic processes. However, in relation to the foot there are two commonly occurring manifestations: distal symmetrical sensorimotor peripheral neuropathy and peripheral sympathetic autonomic neuropathy.

**Distal symmetrical sensorimotor peripheral neuropathy**

Distal symmetrical sensorimotor peripheral neuropathy is characterized by abnormality or loss of sensation in the foot and is broadly divided into painful or painless forms although it is possible to experience both simultaneously. A recent German
population-based study identified a prevalence of neuropathy of 28% in the diabetic population [14].

Sensory thresholds for the development of ulceration are not known and given the complex influence of other complications associated with diabetes, a finite value for ulcer formation may never emerge. Whilst thresholds are used for the purpose of screening assessments, it should be acknowledged that sensation may be present but is no longer functional as a means of protection.

Frequently reported symptoms include unpleasant, abnormal sensations such as pins and needles, tingling, shooting pains, stabbing pains, “like walking on marbles or barefoot on hot sand,” extremes of heat and cold (see Table 66.1). In contrast, patients may experience a feeling of numbness, heaviness or not uncommonly, be completely asymptomatic despite having significant sensory deficits on examination of the lower limb. Some patients may experience a combination of painful and painless symptoms.

Diabetic peripheral neuropathy may alter gait patterns due to impaired sensory feedback (proprioception) during walking [15]. Decreased balance and coordination may have the potential to increase the risk of foot trauma, deformity, and/or falls. All of the above can disturb gait mechanics and a high stepping gait, stamping gait and drop foot gait may be apparent during mobilization.

The motor component of distal neuropathy involves muscle wasting and the smaller muscles of the feet are particularly vulnerable. Claw toes may be observed as a consequence of “unopposed pulley” of the long extensor and flexor tendons due to wasting of the small muscles of the foot. Claw toes and bunions increase the risk of ulceration due to their contribution to creating abnormal pressure distribution on the foot prompting callus formation and a focus for ulceration.

**Peripheral sympathetic autonomic neuropathy**

Autonomic dysfunction in diabetes can affect the cardiovascular system, gastrointestinal tract, urogenital tract, and skin. Anhydrosis in the foot can be identified by dry, cracked skin which can be prone to fissures and ulceration.

Autonomic dysfunction can also disturb precapillary sphincter function causing arteriovenous shunts to open which can give the impression on examination of a well-perfused foot when in reality there is capillary ischemia. Venous pooling and edema may also be present. Thus the warm, dry, insensate foot is very much an “at-risk” foot.

**Other risk factors for foot ulceration**

Ulceration and amputation risk increases with age and duration of diabetes [3,16]. In Western countries, males have been associated with a 1.6-fold increased risk of ulcers and have a higher propensity for amputations compared to their female counterparts [17]. Europeans have a higher risk of ulceration and subsequent amputation than those within the Indian subcontinent Asia or of African Caribbean descent [18].

Patients with a past history of ulceration or amputation are at the highest risk of recurrent ulceration [16]. Poor vision as a result of retinopathy is a predictor of foot ulceration [3]. Peripheral edema causes a restriction in local circulation and has been associated with an increased risk of ulceration [19].

Any alteration in foot architecture exposes the foot to areas of high pressure. Clawing of the lesser digits, hallux valgus deformity, and plantar flexed metatarsal heads with distal migration of the plantar fat pad into the sulcus increase the risk of tissue breakdown in the insensitive foot [3].

The accumulation of callus from peripheral sympathetic dysfunction in neuropathy is highly associated with ulceration. Callus acts as a foreign body within the patients shoe and ultimately leads to tissue breakdown [20].

**Nephropathy**

Renal disease, even in the preliminary stage of microalbuminuria, is associated with an increased risk of foot ulceration [21]. For new dialysis patients, the strongest predictor of lower extremity amputation is diabetes with the incidence in the first year of treatment nearing 6%. There is a strong temporal relationship between commencing dialysis and an increase in the risk of ulceration as demonstrated by Game et al. 2006. This study was retrospective and comparisons were made between patients before and after commencing dialysis. The authors reported a steep increase in the incidence of foot ulceration after initiating renal replacement therapy relative to the pre-dialysis period [22].

The majority of studies reporting an association between nephropathy and diabetic foot ulcers had not made a distinction

### Table 66.1 The variable manifestations of diabetic peripheral neuropathy

<table>
<thead>
<tr>
<th>Stage of neuropathy</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>No neuropathy</td>
<td>No symptoms or signs</td>
</tr>
<tr>
<td>Chronic painful neuropathy</td>
<td>Burning, stabbing pains +/- pins and needles; Absent sensation to several modalities. Reduced/absent reflexes.</td>
</tr>
<tr>
<td>Acute painful</td>
<td>Symptoms as above. Hyperesthesia common. May follow initiation of insulin in poorly controlled diabetes; signs minor or absent.</td>
</tr>
<tr>
<td>Painless with complete/partial sensory loss</td>
<td>Numb/deadness of feet or no symptoms; risk of painless injury, reduced/absent sensation; reduced thermal sensitivity, absent reflexes.</td>
</tr>
<tr>
<td>Late complications</td>
<td>Foot lesions, neuropathic deformity, nontraumatic amputation.</td>
</tr>
</tbody>
</table>
between patients with renal impairment receiving dialysis and those with renal impairment not receiving dialysis. In 2010 dialysis was reported as an independent risk factor for foot ulceration in patients with diabetes and renal impairment [23]. Furthermore, long-term hemodialysis has been demonstrated to undermine the ethnic protection against neuropathy and foot ulcer risk that has previously been indentified in patients with diabetes but not having dialysis [23,24]. The physiology underlying the relationship is yet to be established but could be related to metabolic changes associated with dialysis, long periods lying down or staying in one position whilst undergoing dialysis and the overall shift in disease management towards the renal system possibly to the detriment of other care needs.

Incidence of neuropathic foot ulcers

The relationship between neuropathy and diabetic foot ulceration has been firmly established by decades of research. Data from population- and clinic-based studies concur that the prevalence of neuropathy in patients with diabetes stands at approximately 25–30% [25]. The DCCT Research Group reported an incidence of 7% for those enrolled to receive the intensive treatment regime and 5% for those in the conventional treatment arm. Furthermore, the incidence was seen to increase over time and reassessment at 13–14 years after the DCCT closed reported an incidence of 22% for those in the intensive treatment group and 28% for those in the conventional group [13].

The EURODIAB IDDM Complications Study obtained a baseline prevalence of 28% for neuropathy in T1DM with a quarter of subjects having developed distal peripheral neuropathy over a 7-year period [26].

The North-West Diabetes Foot Care Study calculated an annual incidence of 6% for foot ulceration associated with neuropathy in comparison to those without neuropathy who had an annual incidence of ulceration at 1.1% [3].

Neuropathy is concurrent with 90% of foot ulcers as identified in the North-West Diabetes Foot Care Study whilst a study of the vibration perception threshold as a predictor of ulceration identified that patients with a moderate to severe sensory loss had a sevenfold annual increase of ulceration compared to those with normal sensation [27]. Patients with neuropathic diabetic foot ulcers have a 7% increased risk of amputation in the following 10 years [28].

The pathway to foot ulceration

Diabetic foot ulcers are highly heterogeneous and as such the processes involved in ulcer formation do not lend themselves to a unitary, linear model of pathogenesis. As stated earlier, the main contributors are thought to be DPN, PVD, foot deformity, and altered pressures during gait. The Rothman model of causation has been used to determine the factors involved in ulcer formation and establish causal pathways for patients with diabetes [19]. The most frequently occurring component cause was peripheral neuropathy with foot deformity followed by minor trauma. A component cause alone is insufficient to achieve the endpoint of a DFU, but when combined with other component causes such as minor trauma, a sufficient cause can be generated leading to ulcer development. The model allows for different combinations thereby producing a variety of pathways, for example a patient with altered sensation in the foot and clawed toes has been wearing footwear insufficient to accommodate the foot deformity and presents to the footcare team with an ulcer on the dorsal surface of the foot. The factors involved are component causes as follows: neuropathy + deformity + trauma. Minor trauma was identified as the only factor with the potential to act as a sufficient cause for ulceration, that is, could act alone.

Abnormalities in foot pressures and their distribution may also contribute to ulcer development especially involving the plantar surface of the foot. Extremely high pressures during ambulation can arise from intrinsic, extrinsic or behavioral factors such as alterations in foot structure, including soft tissues, inappropriate footwear, trauma, or walking barefoot. Elevated pressures are most commonly recorded for the forefoot although the mid-foot and heel can be affected.

Patients from other clinical groups such as rheumatoid arthritis also demonstrate abnormal plantar foot pressures during mobilization related to bony prominences but do not have a tendency to ulcerate. This is assumed to be due to their ability to modify their gait in order to avoid loading a vulnerable area of the foot [29]. The patient with diabetes and peripheral neuropathy has the disadvantage of having lost the protective sensation that would prompt a change in gait or choice of softer footwear. Continued weight bearing and exposure to repetitive high pressures can therefore lead to ulceration.

The identification of causative pathways leading to ulceration in the foot highlights the areas in which to target prevention strategies.

Identification of the high-risk foot

The patient with diabetic neuropathy has lost the mechanisms that connect the individual and their feet with the environment. In the absence of this close relationship the foot becomes vulnerable necessitating a shift to the visual system as a means of monitoring integrity of the foot. This task can be performed by healthcare staff during annual reviews but the patient also has a responsibility to regularly monitor their own feet if able to do so. Guidance from the American Diabetes Association document on the “Comprehensive Diabetic Foot Examination” provides clarity on the structure and content of a robust assessment [30].

Assessment

The American Diabetes Association reported on the “Comprehensive Diabetic Foot Examination”. This includes careful
history taking and a clinical examination consisting of a neurologic and vascular assessment.

History
A careful foot examination is the key component of the diabetic foot check.
- History of past or present neuropathic symptoms?
- History of any lower extremity vascular problem (intermittent claudication/rest pain/past history of bypass surgery or angioplasty).
- Past history of ulcer or minor/major amputation?
- Social factors (living alone, blood glucose control, cigarette smoking?).
- Other diabetic complications—especially visual impairment or end-stage renal failure (on dialysis or post-transplant).

Clinical examination
- Skin status — color, thickness, callus, dryness, cracking?
- Any signs of bacterial/fungal infection between toes.
- Breach of the skin/ulceration?
- Alteration in foot architecture, that is, claw toes, prominent metatarsal heads, loss of protective tissue on the sole of the foot.
- Reduced sweating.
- Small muscle wasting resulting in a high arch “intrinsic minus.”
- Skin temperature? A unilateral warm swollen foot with intact skin consider acute Charcot neuroarthropathy until proven otherwise.
- Footwear suitability.

Neurologic assessment
Assessment for loss of protective sensation necessitates the use of two simple tests. In the first instance detection of pressure perception should be assessed using the 10g monofilament (Bailey Instruments Ltd, Manchester, UK) which has been shown in several prospective studies to be a useful predictor of foot ulceration. Sites to be tested are the first, third, and fifth metatarsal heads and the plantar surface of the distal hallux. Failure to detect the perception of pressure at one or more sites in each foot would be considered to be an abnormal response [30].

The result of the monofilament test of pressure perception requires confirmation by a second perception test using one of the following:
- Vibrating 128 Hz tuning fork: this should be tested over the apex of the hallux bilaterally.
- Pin-prick sensation: the inability of a patient to detect pin-prick sensation can be tested using a disposal pin again over the apex of the halluces.
- Ankle reflexes: absence of ankle reflexes in either leg would be regarded as an abnormal response.

Vascular assessment
- Palpation of the posterior tibial and dorsalis pedis pulses: describe as either “present or absent.”
- Doppler ultrasound probe can be useful to assess waveforms although vessel wall calcification can lead to a falsely elevated reading of the ankle-brachial index.

Infection and diabetic foot ulcers
Diabetes-related foot ulcers frequently become infected and the lifetime risk of developing a diabetic foot infection is estimated to be between 4% [31] and 7% [32] but in reality it may be higher. The unfortunate fact is the majority of amputations in patients with diabetes are preceded by an infected DFU [31] and may therefore be avoidable with correct diagnosis and management.

The majority of diabetic foot infections are highly likely to be caused by Gram-positive organisms. *Staphylococcus aureus* is the most common isolate although beta-hemolytic streptococci may also be present. Deep wounds, however, are largely polymicrobial in nature harboring a wide range of bacteria. Microorganisms such as gram-positive cocci, gram-negative rods, and obligate anaerobes may all be present within a single foot wound [33].

Diagnosis of infection in a diabetic foot wound can be problematic due to reduced clinical signs of the host inflammatory response particularly associated with neuropathy and ischemia. There is also the possibility of impaired immune responses in patients with diabetes as illustrated by impaired phagocytosis, reduced intracellular killing and increases in pro-inflammatory cytokines [31].

Whilst some patients with infected wounds will demonstrate pain, warmth erythema, raised temperature or raised CRP, approximately 50% may be asymptomatic (see Figure 66.1). Consideration of more subtle signs of infection may be necessary such as any unexpected changes in wound characteristics, new onset of tenderness, prolonged healing, and wound malodor.

Figure 66.1 Infected diabetic foot lesion. A 56-year-old patient presented with a painless, necrotic ulcer on the 5th toe of the left foot. On examination there was surrounding erythema and purulent discharge clinically suggesting infection. (For a color version of this figure, please see color plate section.)
alterations in discharge characteristics or poor granulation tissue. Moreover, some patients may present with unexpectedly poor glycemic control as assessed by home blood glucose monitoring, unaware that the cause is an infected neuropathic or neuroischemic foot ulcer.

Infection is suspected based on clinical assessment of the wound characteristics and if so antibiotic treatment should be commenced in addition to ongoing wound management. A tissue sample or deep wound swab should be taken for culture and sensitivity to inform a specific antibiotic regimen. Superficial cultures are inappropriate for infected DFUs due to probable contamination by colonizing bacteria during sample collection and therefore tissue cultures are advised.

Suitable antibiotics should include cover for both soft tissue and bone, for example flucloxacillin, cephalixin, and clindamycin, cover both bone and soft tissue, respectively. Sharp debridement, drainage of purulent discharge and off-loading form an essential part of the treatment combination.

Moderate to severe infection can be accompanied by systemic signs such as fever, tachycardia, pyrexia, uncontrollable hyperglycemia plus altered wound characteristics including additional cellulitis, gangrene, and abscess or wound probing to bone. Depending on the severity of the clinical presentation, admission to hospital for parenteral antibiotic administration may be necessary. Surgical debridement or resection of the infected structures may also be indicated.

Management of diabetic foot ulcers

Wound closure is the ultimate aim in the treatment of DFUs. The key elements of intervention include: removal of pressure, restoration of perfusion, eradication of infection if present and local wound care. These goals may be achieved through a combination of different techniques which are summarized in the following (see Figure 66.2).

Debridement

Hyperkeratotic tissue develops on the plantar surface of the foot as a result of shear pressure. Regular sharp debridement of excess keratin that forms callus provides a reduction in abnormally elevated plantar pressures and in doing so reduces the risk of ulceration [34]. Removing this overlying layer, and any necrotic tissue underneath, allows the ulcer to drain any fluid it contains and to begin to heal naturally from the base.

Wounds extending to bone and infected soft tissues require deeper and aggressive debridement to achieve removal of nonviable tissue and provide drainage of purulent discharge. Complete excision can significantly reduce the number of days taken to heal compared with ulcers managed conservatively. Strict off-loading is necessary after aggressive debridement.

Off-loading

Off-loading should be the first-choice treatment option for plantar foot neuropathic ulcers [35]. The aim of off-loading of neuropathic and noninfected neuroischemic plantar foot ulcers is to reduce pressure at the ulcer site through decreased weight bearing and redistribution of pressure over the wider area of the plantar aspect of the foot. Devices available include removable or fixed casts, orthotic devices, and custom-fabricated shoes.

The “total contact cast” (TCC) is the most effective off-loading device and has become the gold standard. Healing times are significantly shorter when total contact casting is used as opposed to other methods of off-loading. The success of total casting lies with the fact that they are nonremovable therefore the patient is obliged to wear them at all times. A study examining the activity patterns of patients with removable casts illustrated poor compliance with treatment and subjects wore the casts for only a minority of steps per day [36]. Thus the use of removable cast walkers rendered irremovable has been proposed for some cases [37]. In some specialist centers, a removable Scotchcast boot has also been used with some success. The boot gives support up to the level of the ankle, can be removed at night and the patient is able to continue mobilizing using a specialist sandal.

Wound dressings

Wound healing is significantly impaired amongst the diabetic population and further complicated by neuropathy and/or ischemia. The use of specialist dressings can provide favorable conditions for healing to occur by providing a moist protective environment. The range of dressings available is vast despite a questionable evidence base.

The basic requirements for any wound dressing product are effective absorption of exudate, thermal insulation, gas permeability and impenetrable to microorganisms. A product should not adhere to the wound itself thus preventing inadvertent removal of newly granulated tissue but at the same time material should be easily removed to allow visual monitoring of the wound it is protecting. Selection of the ideal dressing will depend upon the specific characteristics of the ulcer. Cleansing, removal
Table 66.2 Wound management products

<table>
<thead>
<tr>
<th>Dressing</th>
<th>Description</th>
<th>Contraindications</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hydrocolloid</td>
<td>Facilitate re-hydration and autolytic debridement. Dry, sloughy, necrotic wounds. Promote granulation.</td>
<td>Infected wounds. Twice weekly change.</td>
<td>Aquacel (ConvaTec) Comfeel (Coloplast)</td>
</tr>
<tr>
<td>Hydrogels</td>
<td>Donates liquid to dry wounds and absorbs exudates. Dry, sloughy wounds. Autolytic debridement.</td>
<td>Hydrogel sheets avoided in infected wounds</td>
<td>Intrasite gel (S&amp;N Hlth.) Iodosorb (S&amp;N Hlth.)</td>
</tr>
<tr>
<td>Silver</td>
<td>Antimicrobial. Colonization.</td>
<td>Sensitivity to silver</td>
<td>Acticoat (S&amp;N Hlth.)</td>
</tr>
<tr>
<td>Vapor-permeable</td>
<td>Provide a moist healing environment. Mild exude.</td>
<td>Heavily exudating wound</td>
<td>Tegaderm (3M)</td>
</tr>
<tr>
<td>Foam dressing</td>
<td>Primary or secondary cover. Light and heavy exudates.</td>
<td>Remove if strike through occurs</td>
<td>Allevyn (S&amp;N Hlth.) Lyfofoam (Medlock Medical)</td>
</tr>
<tr>
<td>Larval therapy</td>
<td>Debridement, promote granulation. Heavily sloughy necrotic wounds.</td>
<td>Increase in pain</td>
<td>Maggots (Zoobiotic)</td>
</tr>
<tr>
<td>Alginates</td>
<td>Hemostat. Heavy exudates.</td>
<td>Blockage. Loose fibers.</td>
<td>Kaltostate (ConvaTec)</td>
</tr>
<tr>
<td>Skin substitutes</td>
<td>Living skin. Obstructive wounds.</td>
<td>Colonized. Infected wound.</td>
<td>Dermagraft (S&amp;N Hlth.)</td>
</tr>
<tr>
<td>Iodine</td>
<td>Antibacterial. Exudating wounds.</td>
<td>Iodine (sensitivity) Renal/thyroid conditions.</td>
<td>Iodosorb (S&amp;N Hlth.)</td>
</tr>
<tr>
<td>Honey</td>
<td>Antimicrobial. Sloughy necrotic wounds. Autolytic debridement.</td>
<td></td>
<td>Mesitran (Medlock Medical)</td>
</tr>
</tbody>
</table>

of debris, granulation, vascularization and epithelization are the targets of wound management. As the wound progresses through different stages of healing it may be necessary to use a variety of different dressings (see Table 66.2). Products available can be divided into broad categories of those with debriding properties, antiseptic-based dressings, moisture providing and those that influence the healing process itself. A thorough knowledge of the products available and their theoretical attributes is necessary for selection of the most effective dressing.

**Negative pressure wound therapy**

The use of negative pressure wound therapy is becoming more widespread in the form of the V.A.C. system. Studies have demonstrated faster rates of healing of diabetic foot ulcers and more wounds achieving closure [38]. Increased perfusion and promotion of granulation tissue formation are the reported benefits arising from the cell-stretching effect of negative pressure.

**Growth factors and skin substitutes**

Wound healing involves a complex interaction of a number of growth factors, one of which is platelet-derived growth factor (PDGF). There is growing interest in the potential application of growth factors to aid wound healing in DFUs. Becaplermin is recombinant PDGF ointment and its use has shown some slight benefit. Another growth factor is granulocyte colony stimulating factor (GCS-F) and it has been reported to improve resolution of infection in one pilot study whilst another study claimed it reduced amputation rates but further substantiation is required. Bioengineered skin (Apligraf) and human dermis (Dermagraft) are new types of biologically active implants for ulcers and contain human fibroblasts that deliver GFs and ECM components. However, the evidence base for many of these expensive therapies is weak and further large-scale randomized controlled trials are still needed, and should control as best as possible for the many potentially confounding variables, particularly offloading.

**Multidisciplinary team input**

The delivery of care for patients with diabetes-related foot complications has altered over recent years. The emphasis has
transferred from a centralized, core diabetes footcare team to a delegation of roles based in the community. Increased awareness among healthcare professionals and a shift away from hospital-based care has resulted in changes for the footcare team.

Screening for diabetes-related foot ulcers takes place at a community level with opportunities arising in a variety of different environments. The footcare team now extends to GPs, district nurses, practice-based nurses, and community-based podiatrists.

Successful management of diabetic foot complications depends upon achieving stability in all aspects of diabetes care. Strict glycemic control, stable kidney function, a functional visual system, stable blood pressure avoiding hypotensive episodes and intact cognition are all aspects of diabetes that significantly influence the success or failure of interventions for the lower limb. This can only be achieved through multidisciplinary working including across environmental boundaries.

Patients requiring a total package of care from a specialist diabetes footcare team will require a structured management plan in order to contend with the multiple comorbidities and complications associated with diabetes. A specialist diabetes footcare team should consist of a diabetologist, specialist foot surgeon (podiatric or orthopedic surgeon), specialist diabetes nurse, and podiatrist [39].

Improved outcomes, including reduced incidence of minor and major amputations have been demonstrated in a number of studies when care is delivered in this way. A recent study directly compared outcomes associated with care delivered by an established multidisciplinary diabetes team (MDT) with another hospital lacking a designated diabetes team. Results showed a significant reduction in major amputations performed on patients treated by the diabetes MDT (4.7%) versus 21.7% without MDT input (p < 0.0001). Mortality during hospitalization was also significantly different between the two groups at 2.5% for the MDT group and 9.4% for the controls (p < 0.001) [40].

Advice regarding glycemic control and maintenance of stable glucose levels will assist in achieving overall improvements in diabetes complications. Regular checks for retinopathy will reduce the risk of foot trauma occurring due to visual impairment.

**Patient education**

Patient education has long been advocated as a means of increasing patient understanding of their condition thereby increasing compliance and a subsequent reduction in diabetic foot complications (see Figure 66.3). Guidelines regarding the content and delivery of patient education programs have been included in clinical guidelines from the American Diabetes Association and International Diabetes Federation [41,42].

Comprehension of the features of peripheral neuropathy and its implications can be difficult for patients to accept but without such acceptance daily footcare is unlikely to assume the important position it requires. Successful self-management requires motivation and compliance from the patient to accept a degree of responsibility for their own care; however, foot inspection can be problematic for obese individuals and those with visual impairment. Lavery et al. compared outcomes between patients given standard footcare review and advice, patients instructed in daily structured footcare examination versus patients taking daily temperature readings observing for significant increases as a prompt to contact healthcare services. Although compliance was poor across all the groups, there was a > fourfold decrease in the risk of developing a foot ulcer for patients taking their foot temperatures compared with standard foot management: if a significant difference between the foot temperatures was observed, patients were instructed to rest and contact their podiatrist for further management (OR 4.48; p = 0.008). Providing patients with an objective means of self-monitoring appears to provide a valuable method for encouraging ulcer prevention [43].

The use of inappropriate footwear, both incorrect size and those without inadequate cushioning are known to play an important role in the development of ulcers in patients with neuropathy. Tight shoes commonly lead to ulceration at dorsal deformities such as bunions or between the spaces of toes which have been crushed together. However, loose shoes can also lead to ulceration from the foot slipping inside, creating frictional force. Cushioning refers to the thickness of the soft material under the foot. Even simple sports trainers can reduce planter pressures by 50% compared to leather soles.

Additionally patients should be advised about other associated risk factors such as controlling high blood pressure, cholesterol, smoking cessation, and obesity. Not only will these measures
reduce patients’ risk of ulcers, it will also lower their macrovascular complication risk.

**Charcot neuroarthropathy**

This condition is inextricably linked with distal symmetrical neuropathy as a result of diabetes although the exact mechanisms underlying the progression from neuropathy to tangible changes in foot architecture remain unclear. A number of hypotheses exist but evidence is often lacking.

With regard to incidence data is limited; one study that followed up patients with CN over 15 months reported incidence to be 0.3% per year [44]. An American study of diabetes complications in 2003 identified a total incidence of Charcot deformity to be 0.3% per year [45]. Information on prevalence is sparse; however, this may well be due to the fact that it is notoriously difficult to diagnose, there are no standardized criteria against which symptoms can be compared, and as a result CN can be overlooked as a possibility.

CN is characterized by osseous and joint destruction. Abnormalities may occur in the forefoot, midfoot, peritalar and ankle regions whilst avulsion fractures can affect the posterior tuberosity of the calcaneus. Patients have no awareness of the underlying changes in foot structure due to poor sensation therefore continue to mobilize. Foot deformity associated with CN is determined by the anatomical regions involved in the degenerative process. Changes in the architecture of the mid-foot can lead to collapse of the medial column. Hindfoot involvement can result in dislocation and fragmentation of the transverse tarsal joints and subtalar joint. Ankle joint fractures are less common than the former and latter regions but a more rapid progression in bone destruction results in poorer outcomes. There is a risk of ulceration as a result of increased and abnormally distributed foot pressures.

Although it was first diagnosed approximately 180 years ago, little is known regarding pathogenesis; however, there is general acknowledgment among medical professionals that there are four stages of the condition: a prodromal inflammatory phase when clinical and radiographic symptoms are minimal; a development phase with clinical and radiographic symptoms; a coalescence phase whereby clinical symptoms are marginally reduced but X-rays illustrate a wide range of degenerative osseous changes associated with advanced disease. The final remodeling phase can involve minimal clinical symptoms but fixed, residual deformity may be present whilst radiographically there is evidence of remodeled and new bone formation [46]. The process driving the change is uncontrolled inflammation and it is proposed that the initial inflammatory response is exacerbated by continued mobilization on the fractured joint. The subsequent prolonged production of inflammatory cytokines, RANKL, NF-κβ, and osteoclasts may be the route to ongoing osteolysis. Vascular calcification is also associated with CN and a recent study confirmed a RANKL signaling pathway as a possible signaling pathway mediating this calcification. The study was in vitro and results appear to provide a link between vascular and bone metabolism [47].

**Diagnosis of Charcot neuroarthropathy**

Difficulties in diagnosing CN arise from ambiguous symptoms in the early stages for the condition. Acute local inflammation can be an early sign of underlying joint injury; however, this can also be indicative of cellulites, deep vein thrombosis, and acute gout. A temperature differential between the two feet of several degrees is likely whilst pedal pulses are usually bounding, all of which is characteristic of intact or even exaggerated blood flow to the foot. Plain radiographs can provide information on bone structure, alignment and mineralization. In the early stages of the disease subtle fractures may not be visible but as symptoms progress fractures, dislocation and subluxations may be more overt. Although plain X-ray films are valuable in working towards a diagnosis of CN the disadvantage is the delay between CN changes and visibility on X-ray film. Magnetic resonance imaging (MRI) can capture the early onset changes that may not be visible on plain X-ray films and is already known to have high sensitivity and specificity for osteomyelitis [48]. Positron emission tomography (PET) scanning is a possible modality for future imaging of complex foot pathologies in diabetes and may also be able to distinguish between osteomyelitis and CN.

Above all, the selection of investigations should be underpinned by a strong clinical suspicion of CN as a possibility [49].

**Medical management of Charcot neuroarthropathy**

Total immobilization and off-loading of the foot is the mainstay of treatment for acute, active CN and should be achieved through total contact casts. Use of crutches or a wheelchair can protect the foot further and avoidance of weight-bearing on the affected side is vital. Regular cast changes are necessary to avoid pistoning (movement within the cast) as edema subsides. Another offloading device is the removable cast walker although this will increase the risk of weight bearing on the affected limb. A stepwise return to normal footwear can be guided by improvements noted on imaging studies and reduction in clinical symptoms such as decreased edema, erythema, and improved skin temperatures [50].

Antiresorptive medications have been investigated as a possible treatment based on the high bone turnover that is associated with active CN. Bisphosphonates such as pamidronate and alendronate have been used with the aim of reducing bone turnover and although results were promising, sample sizes were small [51,52]. Large randomized controlled trials are required.
to confirm efficacy prior to using pharmacotherapy in the treatment of CN [46].

**Surgical management of Charcot neuroarthropathy**

Surgical management of CN is based on expert opinion and case series. Rationale for surgery can include bone resection for osteomyelitis, removal of osseous prominences and foot deformities that cannot be managed through bespoke footwear. Tendon lengthening has been used to decrease the CN mid-foot deformity and following total contact casting favorable results are reported [53]. Reconstruction of deformities with internal fixation has also been reported but outcomes have been less positive given the poor bone stock and attenuated soft tissue [54].

**Conclusion**

Diabetic foot complications can arise as a result of the complex interaction between microvascular and macrovascular abnormalities. Peripheral neuropathy plays a major role in the development of foot lesions through loss of sensation. The absence of pain makes the foot highly vulnerable to intrinsic and extrinsic influences which can rapidly steer the patient down the pathway to ulceration.

Guidelines for the prevention and treatment of diabetic foot complications are invaluable in creating the standards of care necessary to achieve a reduction in lower extremity amputations globally.

**References**

CHAPTER 67

Erectile dysfunction in diabetes mellitus

Michael H. Cummings and Rajesh Rajendran
Academic Department of Diabetes and Endocrinology, Queen Alexandra Hospitals Portsmouth NHS Trust, Portsmouth, UK

Key points

• Erectile dysfunction (ED) is extremely common in men with diabetes.
• Endothelial dysfunction and chronic inflammation have been implicated in both ED and cardiovascular disease (CVD).
• Diabetic men presenting with ED should be screened for hypogonadism and CVD.
• Phosphodiesterase type 5 inhibitors (PDE5i) remain first-line agents in the management of diabetic men with ED.
• Hypogonadism may occur in up to one third of diabetic men and testosterone replacement therapy can successfully restore potency in this group.

Introduction

Erectile dysfunction (ED) has been defined as the inability to achieve, maintain or sustain an erection firm enough for sexual intercourse [1]. There have been considerable advances in our understanding of the prevalence, pathogenesis and management of ED in the last few decades. It is now recognized that ED is perhaps the most common complication in diabetic men. There is also a greater understanding of the biochemical pathways that lead to vasodilatation and smooth muscle relaxation within the penile muscle facilitating tumescence. Abnormalities of these biochemical pathways have been identified in diabetic men with ED and have led to the development of a number of successful therapeutic options to treat the condition. In particular, the development of oral phosphodiesterase type 5 inhibitors (PDE5i) has revolutionized management not only because of ease of treatment but also the publicity around their launch has led to a wider understanding and more open discussion of the subject by patients and healthcare professionals. Prior to the launch of this class of drugs ED was considered a taboo subject and both diabetic patients and their healthcare professionals were embarrassed to talk about the problem [2].

Previously, our understanding of ED in diabetic men has been hampered by studies which predominantly focused upon the general population with this condition. The knowledge that diabetes is the commonest cause of ED has led to a barrage of diabetes-specific studies being undertaken and this evidence forms the basis of this chapter examining epidemiology, pathophysiology, and clinical management.

Epidemiology

Despite the variation in populations studied and differing methodologies used to define ED, the overall prevalence of this condition is universally consistent, lying between 30 and 50% [2–7] in diabetic men and is far higher than in nondiabetic men where this figure does not exceed 20% [8]. In most studies, there has been little attempt to segregate findings in type 1 and type 2 diabetic subjects, however a few studies reported a similar prevalence of ED in both forms of diabetes [6,7]. Up to 40% of patients with ED may have underlying diabetes [9] and the incidence of new cases of diabetes has been shown to be around 11–12% [10,11] justifying the screening for diabetes in all previously undiagnosed patients presenting with ED.

Figure 67.1 demonstrates the exponential rise in ED with age and that the prevalence is always higher in the diabetic male compared with the nondiabetic population. The likelihood of ED in a 35-year-old diabetic male is equivalent to that of a 60-year-old nondiabetic male.

McCulloch et al. [4] showed that in diabetic men with ED, the most significant associations with its development were age ($p < 0.001$), treatment with oral hypoglycemic agents or insulin ($p < 0.001$), retinopathy ($p < 0.001$), symptomatic peripheral and symptomatic autonomic neuropathy ($p < 0.001$ and $p < 0.005$, respectively). There was also a weaker association...
with duration of diabetes, poor glycemic control, ischemic heart disease, and nephropathy. This cohort was re-examined 5 years later at which stage only 9% had regained potency [12], reflecting the progressive and nonreversible nature of the condition supporting the concept that the etiology is primarily organic in most cases.

In practical terms when attempting to forecast the likelihood of ED in diabetic men presenting for routine review, the probabilities are less than 5% in a 20-year-old male with no diabetic complications, 40–80% in a 40-year-old male with diabetic complications, rising to 80–100% in a 65-year-old male with diabetic complications.

**Pathophysiology**

Table 67.1 shows the factors that may contribute to the development of ED in the diabetic male. The most influential factors are considered to be local changes within the corpus cavernosum and aberrant neurovascular supply to this organ.

**Changes within the corpus cavernosum**

It is now recognized from studies of diabetic animal and human corpus cavernosum that there are abnormalities in the availability or effectiveness of several key chemical mediators involved in smooth muscle relaxation and constriction (Table 67.2), and that ED is strongly associated with endothelial dysfunction.

Nitric oxide (NO) is released either directly from the parasympathetic nerve terminal (PNT), or from the local vascular endothelium following stimulation by PNT released acetylcholine (ACh) and is integral, via its messenger cyclic GMP, in reducing intracellular calcium concentrations and facilitating smooth muscle relaxation. It has been demonstrated in diabetic men with ED that there is impairment in both parasympathetic nerve and endothelium-dependent mechanisms that mediate smooth muscle relaxation [13–15]. Diabetes impairs the activity of guanylyl cyclase, reduces NO synthase levels and increases superoxide radicals, all leading to oxidative stress and free-radical damage finally leading to endothelial toxicity [16] and atherosclerosis. ACh administration was associated with a lesser degree of muscle relaxation in corpus cavernosum from diabetic men with ED and the degree of impairment was associated with the duration of diabetes [13]. Experimental studies in Wistar rats have provided a mechanism linking chronic dysglycemia and ED. When glycated hemoglobin was added to isolated corpus cavernosum from this animal model, ACh-mediated relaxation of this smooth muscle was impaired [17]. Moreover, this effect of glycated hemoglobin was mimicked by the addition of pyrogallol (a donor of superoxide anions) and partially reversed by the addition of L-arginine (the precursor of NO) [18]. In summary, these studies support the concept that dysglycemia impairs NO-induced smooth muscle relaxation and vasodilatation in the diabetic male with ED.

The PNT within the corpus cavernosum also releases another smooth muscle relaxant, vasoactive intestinal polypeptide (VIP), capable of reducing intracellular calcium via its messenger, cyclic AMP. It has been shown that VIP immunoreactivity is reduced in penile tissue of streptozotocin-induced diabetic rats [19] and diabetic men with ED [14]. It is possible that vasoactive prostanooids synthesized and released by the corpus cavernosum may also affect smooth muscle relaxation in the diabetic corpus cavernosum although addition of indomethacin,
a cyclo-oxygenase inhibitor, had no effect upon the glycated hemoglobin mediated impaired cavernosal muscle response to ACh in rats [17].

Noradrenaline, the principal mediator of smooth muscle contraction within the penis appears to be less abundant in the corpus cavernosum of diabetic men with ED [15] although the responsiveness of smooth muscle to noradrenaline seems to be unaffected [14]. Endothelin-1-induced smooth muscle contraction also appears unaltered in corporal smooth muscle strips of diabetic men with ED [20], though it has been shown that basal concentration of endothelin-1 is increased in men with ED and more so in subjects with diabetes [21].

Advanced glycation end-products (AGEs) that result from non-enzymatic reactions between glucose and lipids, proteins or nucleic acids is found in greater quantity in the corpus cavernosum of diabetic patients. Their covalent bonding with vascular collagen increases vascular permeability, reactive oxygen species, endothelial expression of adhesion molecules and procoagulant expression thus resulting in oxidative cell damage, decreased NO, decreased cGMP and impaired cavernosal smooth muscle relaxation [22].

Structural abnormalities of diabetic penile tissue have also been identified. These include fibrosis of the penile arteries, corpus cavernosum and reduced amounts of cavernous nerve fibers of unmyelinated axons with collagen, factors which will impair smooth muscle relaxation and vasodilatation. Altered gene expression (downregulated mRNA transcript) [23] and increased apoptosis [24] have been noted in penile tissue from diabetic men with ED and rats, respectively.

**Neurologic factors**

Neuropathy, particularly autonomic neuropathy, has been linked to the development of ED [25], especially under the age of 60 and the principal abnormality appears to be within the parasympathetic nervous system, responsible for achieving tumescence. This is evident from clinical, observational [4,12], and physiologic studies. In some diabetic men, dysfunction of the penile nerves precedes neuropathy in the other peripheral nerves [26]. Abnormalities in deep breathing tests [27], thermal sensation [28], and urinary flow rates [29] that are linked to the integrity of the parasympathetic nervous system are all impaired in diabetic males with ED. By contrast there appears to be no impairment of the sympathetic nervous system [27]. Perhaps surprisingly, despite the clinical link between ED and a symptomatic sensory neuropathy [4], physiological studies undertaken using several different tools have largely shown no difference in sensory innervations between diabetic men with ED and potent males [29,30].

At the cellular level, neurologic changes in diabetic men with ED may be due to hyperglycemia-induced activation of protein kinase C, aldose reductase pathways and non-enzymatic glycation which provokes oxidative stress [31].

**Vascular factors**

The penis is a highly vascular organ and diabetes may be associated with diffuse atherosclerosis or isolated disease of the external iliac artery [32] both of which may contribute to the development of ED. Observational studies have linked the presence of ED with complications in other microvascular and macrovascular beds of men with diabetes [4,12]. It has been shown that significant ED is associated with diabetic retinopathy severity that is independent of age, duration of diabetes, and cardiovascular risk factors [33], and that ED in diabetic men is associated with proteinuria which is a marker for diabetic nephropathy [34]. Diabetic neuropathy is integrally linked to impaired microvascular blood supply. Abnormalities of vascular function, in particular NO-mediated vasodilatation have been described earlier (see section “Changes within the corpus cavernosum”). Up to 95% of diabetic men with ED show impaired penile blood flow [35] and this has been suggested to be more likely as an etiologic factor in type 1 compared with type 2 diabetic men with ED [36].

**Hypogonadism**

Hypogonadism is defined as a clinical condition comprising both symptoms with or without signs and biochemical evidence of testosterone deficiency [37]. It was well known that testosterone regulates male sexual behavior and attitudes but it is now more clear that penile erection, and penile blood flow is associated with circulating testosterone levels and experimental studies in humans and animals have shown that testosterone controls several mechanisms (Figure 67.2) that lead to erection and detumescence. The main physiologic action of testosterone appears to be the regulation of timing of the erectile process as a function of sexual desire [38].

The link between diabetes, hypogonadism and ED has become better recognized in the last decade [37–39]. Several cross-sectional studies including those measuring free testosterone levels have shown the prevalence of biochemical and symptomatic hypogonadism in diabetic men from 33–42% [37], justifying the screening for hypogonadism in this population. Though traditionally hypogonadism has been classified into primary (low testosterone, raised gonadotrophins) and secondary hypogonadism (low testosterone, low gonadotrophins), more recently it has become established that hypogonadism with low testosterone and normal gonadotrophins (mixed hypogonadism) can occur. This type is usually associated with obesity and aging but it has been noted in nearly two-thirds of hypogonadal men with T2DM [37]. Visceral obesity and metabolic syndrome can directly impact on testosterone levels and vice versa. Yet, the mechanism of hypogonadism in diabetes is not fully clear. However, a number of mechanisms have been suggested including low circulating levels of plasma sex-hormone-binding globulin causing low total testosterone in diabetes leading to increased insulin resistance, increased aromatase activity resulting in increased conversion of testosterone to estrogen in visceral adipose tissue thereby decreasing testosterone levels, and insulin resistance in T2DM per se leading to reduced insulin action in the hypothalamus resulting in secondary hypogonadism [37–39].
Figure 67.2 The putative role of testosterone in the mechanism of penile flaccidity and erection. (a) Smooth muscle cell contraction in the corpora cavernosa. NE binding to α1 receptors generates IP$_3$, which, by increasing intracellular calcium (Ca$_{2+}$) levels, activates Ca$_{2+}$-sensitive CLCAs resulting in membrane depolarization, with the diffusion of the stimulus to the neighboring cells and the opening of VOC. The increased Ca$_{2+}$ flow promotes, through calmodulin, activation of MLC kinase and cell contraction. Cell contraction is also obtained by altering the Ca$_{2+}$ sensitivity through a NE-induced activation of a second pathway, rhoA/ROCK, which increases, through a series of kinase activation, the sensitivity of MLC to Ca$_{2+}$. Testosterone is proposed to negatively regulate this second pathway. (b) Smooth muscle cell relaxation in the corpora cavernosa. NO is generated by NO synthases in either NANC neurons (nNOS) or endothelial cells (eNOS). Both steps are positively regulated by testosterone. NO diffuses into smooth muscle cells and activates an sGC, which in turn transforms GTP into cGMP. cGMP activates PKG, which, through various pathways, ultimately decreases intracellular Ca$_{2+}$ levels, leading to relaxation. PDE5 metabolizes cGMP into GMP thereby limiting its effects. This event is positively controlled by testosterone. cGMP cyclic GMP; CLCA, Ca$_{2+}$-sensitive chloride channel; CPI-17, protein phosphatase 1 regulatory subunit 14A; eNOS, endothelial nitric oxide synthase; GAP, GAP, GTPase activating protein; GEF, guanine nucleotide exchange factor; IP$_3$, inositol 1,4,5-trisphosphate; MLC, myosin light chain; MLCK, myosin light chain kinase; MLCP myosin light chain phosphatase; NANC, nonadrenergic-noncholinergic; NE, norepinephrine; nNOS, neuronal nitric oxide synthase; NO, nitric oxide; PDE5, phosphodiesterase type 5; PKG, protein kinase G; PLC, phospholipase C; rhoA, ras homolog gene family member A; rOCC, rho-associated, coiled-coil containing protein kinase; sGC, soluble guanylate cyclase; VOC, voltage-operated channels. Source: Corona and Maggi [38]. Corona 2010 [38]. Reproduced with permission of Nature Publishing Group.
Other endocrine factors
It is well recognized that hyperprolactinemia, thyrotoxicosis, hypothyroidism or growth hormone deficiency can be associated with the development of ED. There is no significant difference in the circulating levels of these hormones in the diabetic population with and without ED compared with the general population whether examining random measurements or during dynamic testing [40,41]. Whilst mean concentrations of TSH, thyroxine and tri-iodothyronine do not differ in the diabetic and nondiabetic populations who are either potent or have ED, it is recognized that thyroid abnormalities are more commonly observed in diabetes [42].

Drugs
There has been an exponential escalation in the use of polypharmacy regimens in the diabetic individual. Many of these drugs may be associated with the development of ED and it is not surprising that iatrogenic causes of ED may be on the increase. Whilst for many drugs (e.g. antihypertensive agents) the mechanism responsible for the development of ED is clear, for many other drugs the etiologic process is obscure. Although ED is more common in patients on oral hypoglycemic agents or insulin [4], this is more likely to be due to disease duration and hyperglycemia rather than the treatment per se.

Table 67.3 highlights drugs that may be commonly used in treating diabetes and are associated with the development of ED and also drugs used in the general population that commonly precipitate ED.

Psychological factors
There has been a diverse reporting of the prevalence of psychological or psychiatric disease in diabetic men with ED, ranging between 9 and 39% [10,36,43], but it is not possible to categorically implicate these factors as the primary etiologic precipitant of ED. It has been noted that diabetes doubles the chance of depression [44] and there is a strong positive correlation between depressive symptoms and the incidence of ED among diabetic men [45]. It is generally held that organic disease predominates although performance anxiety and other psychological factors [45] that are common in patients with ED (particularly those with other vascular complications) may perpetuate the problem. Libido, the degree of coital, masturbatory and sleep-related erections, and frequency of sexual satisfaction and coital activity is also reduced in diabetic men [46].

Lifestyle factors
There is accumulating evidence linking a number of lifestyle factors, such as sedentary lifestyle, smoking, obesity and alcohol consumption, to the occurrence of ED [47]. Sedentary lifestyle has been shown to be associated with higher prevalence of ED in both diabetic and nondiabetic men [48] with increased physical activity conferring protection upon ED [49]. With the increasing prevalence of obesity and diabetes worldwide and epidemiologic studies indicating obesity as a significant independent risk factor for ED, a majority of studies have shown that weight loss can improve erectile function in obese men, though the beneficial effect is less profound in obese diabetic men [50]. Whilst it is well known that smoking is a risk factor for ED [47], cessation of cigarette smoking has been shown to improve ED in a considerable proportion of smokers [51].

Other factors
The association between chronic hyperglycemia and ED, and the mechanisms that may be implicated in dysglycemia-induced ED have been discussed. The incidence of ED is increased threefold with an HbA1c level >65 mmol mol⁻¹ [52]. It is also recognized that acute deterioration in glycemic control may result in temporary ED in hitherto potent diabetic males although the mechanism is less clearly understood. Painful balanitis is more commonly observed in diabetic males, particularly those with poor glycemic control. Autonomic dysfunction has been suggested as being causative in penile venous leaks, observed in two-thirds of diabetic males with vascular ED [53]. Structural abnormalities of the penis that may cause ED, for example

Table 67.3 Drugs associated with the development of erectile dysfunction in diabetic men

<table>
<thead>
<tr>
<th>Drugs commonly prescribed in the management of diabetes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiovascular protective drugs</td>
</tr>
<tr>
<td>Thiazide diuretics and spironolactone</td>
</tr>
<tr>
<td>Beta-blockers (propranolol possibly the worst culprit)</td>
</tr>
<tr>
<td>Angiotensin-converting enzyme inhibitors</td>
</tr>
<tr>
<td>Calcium channel blockers and other antihypertensive agents</td>
</tr>
<tr>
<td>Digoxin</td>
</tr>
<tr>
<td>Drugs for symptomatic relief of painful peripheral neuritis</td>
</tr>
<tr>
<td>Tricyclic antidepressant agents</td>
</tr>
<tr>
<td>Carbamazepine</td>
</tr>
<tr>
<td>Gabapentin</td>
</tr>
<tr>
<td>Nonsteroidal anti-inflammatory agents</td>
</tr>
<tr>
<td>Lipid-lowering agents</td>
</tr>
<tr>
<td>HMG CoA reductase inhibitors (statin therapy)</td>
</tr>
<tr>
<td>Fibates</td>
</tr>
<tr>
<td>Drugs of a general nature commonly associated with development of erectile dysfunction</td>
</tr>
<tr>
<td>Alcohol</td>
</tr>
<tr>
<td>Anticonvulsant agents</td>
</tr>
<tr>
<td>Antiarrhythmic agents</td>
</tr>
<tr>
<td>Antipsychotic agents (especially lithium)</td>
</tr>
<tr>
<td>Antidepressant agents (tricyclics and monoamine oxidase inhibitors)</td>
</tr>
<tr>
<td>Anabolic steroids</td>
</tr>
<tr>
<td>Allopurinol</td>
</tr>
<tr>
<td>Acetazolamide</td>
</tr>
<tr>
<td>Baclofen</td>
</tr>
<tr>
<td>Bromocriptine</td>
</tr>
<tr>
<td>Cimetidine</td>
</tr>
<tr>
<td>Estrogen</td>
</tr>
<tr>
<td>ketoconazole</td>
</tr>
<tr>
<td>Metoclopramide</td>
</tr>
<tr>
<td>Opiates</td>
</tr>
</tbody>
</table>


Peyronie’s disease, are no more frequently observed in the diabetic individual compared with the general population.

**ED: a marker for cardiovascular disease (CVD)**

There are two major hypotheses linking both ED and CVD [16]. One of them implicating endothelial dysfunction and chronic inflammation (seen in both ED and CVD) can be caused by various vascular insults such as diabetes, hypertension, smoking, hyperlipidemia, and hypertension. The pathophysiologic mechanisms of endothelial dysfunction at cellular levels have been described earlier (see section “Changes within the corpus cavernosum”). Chronic low-grade inflammation and oxidative stress may lead to atherosclerosis by contributing to all its stages from the initial phase of increased endothelial permeability through to plaque maturation and rupture [16,54]. Increased levels of markers for inflammation have been demonstrated in both CVD and ED in patients with T2DM [54,55].

The second theory is the artery size hypothesis [16]. Due to the smaller diameter of penile arteries, the same level of plaque burden has a greater effect upon blood flow through the penile arteries compared with coronary and carotid arteries and therefore, ED manifests earlier than CVD.

It has now been shown in clinical trials that ED maybe the early clinical manifestation of generalized vascular disease and is an independent risk for cardiovascular events [16]. It can precede the development of clinically evident coronary artery disease (CAD) by a significant period of time [56,57]. The available data also suggests that the severity of ED correlates with the extent of CAD [58]. Two separate prospective studies have suggested that the risk of new onset CVD was higher in diabetic patients who developed ED and longitudinal studies have shown that ED is able to predict future cardiovascular events [16,57]. Therefore, ED could be a potential marker when screening for silent CAD in diabetic patients.

**Management**

The clinical assessment and management of ED in the diabetic male is broadly similar to that of ED in nondiabetic men but a few areas are worthy of particular mention. In addition more modern therapeutic approaches are discussed with emphasis on their effectiveness in the diabetic male.

**Clinical assessment**

**History and examination**

The presence of other microvascular or macrovascular complications may give a clue to the underlying etiology of ED in a given patient. Conversely, it is extremely important to assess for previously unrecognized cardiovascular disease in diabetic men with ED, given the strong associations between this triad and the very high prevalence of CVD. Not only may ED precipitate vascular events in some susceptible diabetic men but it is also likely to influence treatment choice.

The International Index of Erectile Function (IIEF) may be used as a guide of the severity of the ED and response to therapy. The IIEF is a 15-item self-administered questionnaire that was developed as a measure to detect treatment-related erectile function in patients in cross-cultural settings. A more abridged five-item version of the IIEF was then developed (IIEF-5) (Table 67.4), which has been shown to possess favorable properties for detecting the presence and severity of ED. A score of ≤21 may be classified as having ED and >21 as not having ED [59].

**Investigations**

Table 67.5 shows the investigative tools available in clinical practice to assess the etiology of ED in the diabetic male. There is considerable variation worldwide surrounding the extent of routine investigation that should be undertaken and this remains controversial. In our own practice, we restrict investigations to hormonal blood tests although it is recognized that the yield of hitherto unrecognized endocrine abnormalities apart from hypogonadism is low. It is important to screen for hypogonadism as concurrent treatment with both testosterone replacement therapy and PDE5i may be warranted [37]. We avoid extensive investigation unless indicated on a clinical basis since in our experience this rarely influences individual management plans.

**Treatment options**

**General considerations**

Weight loss in obese men, increased physical activity, stopping smoking and limiting alcohol consumption to moderate amounts are some important lifestyle changes that need to be reinforced in patients with ED in T2DM [47,60,61]. Sexual activity per se is equivalent to mild to moderate physical activity in the range of 3 to 5 metabolic equivalents (METS; i.e., the equivalent of climbing two flights of stairs) [61]. This is important to understand as reassurance can be given to patients with CAD who are able to achieve this level of exercise without symptoms. Overall it is extremely rare for sexual activity to trigger a cardiac event (30 chances per million per hour in the 2-hour period following sexual activity). The second Princeton consensus guidelines (Table 67.6) have been extremely useful in assessing the cardiovascular risk of a patient in relation to considering treatment for ED [60].

Whilst chronic hyperglycemia is integrally linked to the development of ED there is no evidence that supports restoration of or enhanced erectile performance if glycemic control is improved. However, recently it has been shown in a small group of patients that kidney and pancreas or kidney transplantation alone was associated with improvement in sexual function suggesting that achievement of normoglycemia was paramount in improving sexual function in this group [62]. ED developing
Table 67.4: The five-item version of the International Index of Erectile Function (IIEF-5) questionnaire

<table>
<thead>
<tr>
<th>Over the past 6 months</th>
<th>Score awarded for each item</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item 1: How do you rate your confidence that you could get and keep an erection?</td>
<td>Very low</td>
</tr>
<tr>
<td>Item 2: When you had erections with sexual stimulation, how often were your erections hard enough for penetration?</td>
<td>Almost never/never</td>
</tr>
<tr>
<td>Item 3: During sexual intercourse, how often were you able to maintain your erection after you had penetrated (entered) your partner?</td>
<td>Almost never/never</td>
</tr>
<tr>
<td>Item 4: During sexual intercourse, how difficult was it to maintain your erection to completion of intercourse?</td>
<td>Extremely difficult</td>
</tr>
<tr>
<td>Item 5: When you attempted sexual intercourse, how often was it satisfactory for you?</td>
<td>Almost never/never</td>
</tr>
</tbody>
</table>

The IIEF-5 score is the sum of the responses to the five items. The score can range from 5 to 25.

Table 67.5: Investigations of erectile dysfunction in diabetic men

- Measurements to assess glycemic control (glycated hemoglobin, fructosamine)
- Hormone measurements (testosterone axis, prolactin, thyroid function tests, growth hormone, IGF-1)
- Prostatic specific antigen measurement
- Penile Doppler studies
- Responsiveness to a standard intracavernosal injection (“papaverine test”)
- Physiologic tests of the autonomic nervous system
- Detailed psychosexual assessment
- Nocturnal penile rigidity studies
- Invasive radiologic procedures (cavernoscopy and arteriography)
- MRI pituitary gland

Testosterone replacement therapy in men with hypogonadism

Testosterone replacement in eugonadal men has not been demonstrated to improve erectile performance and has potentially dangerous long-term side effects such as sequelae of resulting polycythemia. In hypogonadal men, testosterone replacement is a successful and relatively fast-acting therapy with recipients often noticing dramatic improvement within days of commencing treatment. Given the short half-life of oral testosterone rendering it ineffective, most authorities consider replacement by other routes. Intramuscular testosterone (typically given every 3 weeks with standard testosterone or every 3 months with longer acting preparations) is well tolerated although painful injection sites may be problematic. Topical preparations including gels, creams, and patches are also popular although practical considerations (avoiding contact with water for a period after application, the presence of sweat on the skin or rashes with the latter) can limit their use. Other less commonly used routes include buccal testosterone, and subcutaneous implants inserted every 4–5 months. Care should be given to assess response (libido and erections) in light of achieving therapeutic replacement levels of testosterone. Other benefits of testosterone may include improved energy levels, muscle bulk and power, cognitive function, mood, and bone density. It is imperative to assess for side effects such as prostatic disease and polycythemia. In men who previously failed to respond to PDE5i and now diagnosed hypogonadal, it has been

Concomitantly with the onset of acute deterioration in glycemic control has a more positive outcome if good glucose control can be re-established. Although diabetic men with ED are commonly taking drugs associated with the development of ED, it is our experience that restoration of potency is usually only successful in those patients where there has been a temporal relationship between the commencement of a drug (within 2 weeks) and the development of ED. Moreover, even if implicated in the etiology, there is often compelling clinical grounds to continue with the drug. The majority of patients will need to consider the therapeutic options summarized in Table 67.7. Figure 67.3 is a useful algorithm to guide physicians in treating diabetic men with ED.
Table 67.6 The second Princeton consensus guidelines on cardiovascular risk stratification in patients with sexual dysfunction

<table>
<thead>
<tr>
<th>Grading of risk</th>
<th>Cardiovascular status at presentation</th>
<th>Recommendation for the management of ED</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Low risk</strong></td>
<td>Controlled hypertension.</td>
<td>Manage ED within primary care setting.</td>
</tr>
<tr>
<td></td>
<td>Asymptomatic and &lt;3 major risk factors for CAD, excluding gender.</td>
<td>Review treatment options with patient and partner (where possible).</td>
</tr>
<tr>
<td></td>
<td>Mild valvular disease.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Post MI (&gt;6–8 weeks) post successful re-vascularization (3–4 weeks).</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CHF (NYHA class I).</td>
<td></td>
</tr>
<tr>
<td><strong>Intermediate risk</strong></td>
<td>Recent MI (between 2–6 weeks).</td>
<td>Specialized evaluation recommended (e.g. exercise testing or echocardiography).</td>
</tr>
<tr>
<td></td>
<td>Asymptomatic and ≥3 risk factors for CAD, excluding gender.</td>
<td>Place patient in high or low group depending upon outcome of testing.</td>
</tr>
<tr>
<td></td>
<td>CHF (NYHA Class II).</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Noncardiac atherosclerotic sequelae (e.g. CVA).</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moderate stable angina.</td>
<td></td>
</tr>
<tr>
<td><strong>High risk</strong></td>
<td>Unstable or refractory angina.</td>
<td>Refer for specialized cardiac evaluation and management.</td>
</tr>
<tr>
<td></td>
<td>Uncontrolled hypertension.</td>
<td>Treatment for ED to be deferred until cardiac condition stabilized and/or specialist evaluation completed.</td>
</tr>
<tr>
<td></td>
<td>CHF (NYHA Class III, IV).</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Recent MI (within last 2 weeks).</td>
<td></td>
</tr>
<tr>
<td></td>
<td>High-risk arrhythmias.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Obstructive hypertrophic cardiomyopathy.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moderate/severe valvular disease.</td>
<td></td>
</tr>
</tbody>
</table>

Table 67.7 Options used in clinical practice for treating erectile dysfunction in diabetic men

<table>
<thead>
<tr>
<th>Therapy</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>First-line treatment</strong></td>
<td>Oral therapy</td>
</tr>
<tr>
<td></td>
<td>Testosterone replacement therapy</td>
</tr>
<tr>
<td><strong>Second-line treatment</strong></td>
<td>Intracavernosal therapy</td>
</tr>
<tr>
<td></td>
<td>Transurethral therapy</td>
</tr>
<tr>
<td></td>
<td>Vacuum tumescence devices</td>
</tr>
<tr>
<td></td>
<td>Psychosexual therapy</td>
</tr>
<tr>
<td><strong>Third-line treatment</strong></td>
<td>Penile implants</td>
</tr>
<tr>
<td><strong>Others</strong></td>
<td>Corrective surgery</td>
</tr>
</tbody>
</table>

shown that testosterone replacement therapy subsequently makes patients respond to PDE5i in nearly 60% of cases, underlining the importance of screening for hypogonadism in diabetic men with ED [37–39].

**Oral treatment**
The biggest single advance in the treatment of diabetic ED has been the development of oral PDE5i. PDE5 is the isoenzyme responsible for the breakdown of cyclic GMP, the final chemical messenger promoting smooth muscle relaxation within the penis. Thus PDE5i facilitates erections through prolonging the availability of cyclic GMP within the corpus cavernosum. Currently there are four commercial forms of PDE5i available, sildenafil [63], vardenafil [64], tadalafil [65], and avanafil [66] (Table 67.8).

Their effectiveness is in part determined by the severity of ED, the underlying etiology, the clinical characteristics of the diabetic male being treated and the tool used to evaluate response. These factors make it difficult to reliably interpret any clinical differences in effectiveness between these agents. Broadly speaking, however, most clinicians agree that their effectiveness is similar (between 3 and 4 out of 5 patients will respond), a response rate that is less than that of the general population with ED. PDE5i are generally well tolerated with very few patients stopping treatment due to unacceptable side effects. PDE5i have been demonstrated to be a cost-effective approach in the management of ED [67]. Selective CAD patients with stable disease can take PDE5i to treat ED but not when concurrently using nitrates. PDE5i can have positive effects on cardiovascular prognosis and some studies have shown beneficial effect upon
Erectile dysfunction in diabetes mellitus

Figure 67.3 Algorithm for treatment for a diabetic man presenting with erectile dysfunction.

endothelial dysfunction and fewer cardiovascular events [57]. However, more robust randomized controlled trials are needed to establish the long-term cardioprotective effects of PDE5i. The use of sublingual apomorphine and oral yohimbine has now ceased since the development of PDE5i.

Several over-the-counter herbal medicines or Chinese traditional medicines have been marketed for the treatment of ED, but there are published studies on humans only for some of them such as Panax ginseng, Butea superba, and yohimbine, showing efficacy in the treatment of ED but these are not considered robust studies [68–70]. Besides many of these so-called “natural” products contain either potent inhibitors of PDE5 [68] or pharmacologic doses of PDE5i such as sildenafil or tadalafil [70]. This is clearly a concern when marketed freely as it can have potentially fatal interaction with nitrates and therefore improved regulation of natural health products is paramount.

Intracavernosal injections

In the early 1980s, intracavernosal injections with papaverine [71] and phentolamine [72] were reported as being effective in the treatment of ED. There is a high incidence of priapism and other side effects with both drugs and neither have been licensed for this specific indication. Their use has been superseded by a licensed intracavernosal preparation (prostaglandin E1, alprostadil) which can be self-administered directly into the corpus cavernosum (see Figure 67.4). It is the most potent available treatment for ED (approximately 9 out of 10 injections in the general population resulted in successful erections) and has a lower incidence of priapism [73] compared with previous intracavernosal preparations. The occurrence of priapism (erections lasting greater than 6 hours) should be considered a medical emergency and treatment options to facilitate detumescence include leg exercise, locally applied ice-packs, aspiration of blood from the corpus cavernosum
or intracavernosal injection of alpha-adrenergic drugs such as phenylephrine [74]. Corrective surgery is rarely required. Other side-effects include local pain and tingling, bruising, fibrosis, scarring and occasional infection that may limit its use. One of the largest studies of diabetic men with ED has showed that intracavernosal alprostadil was at least as effective in diabetic men compared with the nondiabetic population and no differences in responses were observed between type 1 and type 2 diabetic patients [75].

Transurethral alprostadil
The application of alprostadil in the form of a pellet (see Figure 67.4) (MUSE) in to the urethra has been associated with a 65.9% success rate in achieving erections suitable for sexual intercourse in the general population [76]. Although there are no major prospective diabetes-specific reports of its use, the study mentioned earlier examined a subgroup of diabetic men with ED and reported similar success rates. Penile pain was reported in one third of patients taking this form of alprostadil and other side effects more commonly reported compared with placebo were minor urethral trauma and dizziness although priapism was not reported.

Vacuum devices
Vacuum devices provide a safe alternative strategy to pharmacologic approaches (Figure 67.4). A cylinder with a constriction rubber band at its base is placed over the penis. A manual or battery pump is operated to create a vacuum in the cylinder leading to tumescence. The rubber band is slipped on to the base of the penis and can be left in place for up to 30 minutes. Side effects are rare and include bruising, ejaculatory failure, and penile discomfort. They require a moderate degree of manual dexterity and are usually only suitable for diabetic men in stable relationships. It has been reported that vacuum devices achieve satisfactory erections in nearly 70% of diabetic men [77]. However, nearly 30% of men discontinued use as a result of inadequate rigidity, penile pain, failure to ejaculate, and poor sexual satisfaction.

Psychosexual counseling
Clearly this is the treatment of choice for diabetic men with an overt psychological or psychiatric etiology to their ED. It may also provide an additional approach to patients receiving physical treatment given the performance anxiety seen in many diabetic men with organic ED. There are many different psychological therapies available. One of the most well-known
forms is sensate focusing employing the Masters and Johnson technique [78]. It requires staged exercises broadly divided into three phases: nongenital, genital, and vaginal containment.

Surgery
Penile prostheses may be considered in diabetic patients who have failed to respond to less invasive measures. There are various complexities of prostheses which range from simple malleable devices that produce a permanent semi-rigid erection to more complex inflatable devices which are associated with an implanted reservoir system enabling tumescence or detumescence as required. Mechanical failure is relatively common, particularly with the inflatable devices. Also extrusion of the prosthesis, pain and bruising may occur, but in centers with good experience of inserting these devices, these complications are rare [79].

Corrective surgery may be considered in patients with primary penile abnormalities such as Peyronie’s disease and venous ligation may be undertaken in diabetic patients with established venous leaks. In diabetic patients with impaired penile blood flow, microvascular surgical techniques have proved disappointing and are rarely performed.

Future treatment options
Topical agents either as monotherapy (nitrates) or combination therapy (e.g. aminophylline, isosorbide dinitrates and co-dergocrine [80]) applied to the penis continue to be explored as do other oral agents such as L-arginine. Phase IIb trials with bremelanotide (a synthetic peptide analogue of α-melanocyte stimulating hormone) have demonstrated efficacy in diabetic men but its challenging adverse effect profile (nausea, emesis, and increased blood pressure) may preclude development of this drug for the treatment of ED [81]. The effectiveness of other intracavernosal agents, for example VIP, are also being evaluated. Although controversial, gene therapy through introduction of vectors that promote the synthesis of nitric oxide synthase (responsible for the production of NO) have been shown to be effective in promoting erections in diabetic rats [82]. Novel treatments such as umbilical stem cell therapy have shown promise in ED (for men with T2DM) and may become available in the future for refractory cases [83].

Conclusions
In the 1980s it was common for diabetic patients with ED to suffer in silence. Both patients and healthcare professionals avoided discussing the subject and treatment options were not known about, services to provide treatment were scanty and successful therapeutic treatments were mainly unlicensed and poorly tolerated. Today, many healthcare professionals involved in diabetes care adopt a holistic approach and can successfully treat ED within their own service without the need for referral to specialist services. Previously, the common approach of ushering diabetic patients with ED out of the consultation room with no treatment other than being told “not to worry, it will get better with time” should no longer occur.

References


CHAPTER 68

Periodontal disease and diabetes mellitus

Wenche S. Borgnakke1 and Robert J. Genco2
1Department of Periodontics and Oral Medicine, University of Michigan School of Dentistry, Ann Arbor, MI, USA
2Department of Oral Biology, State University of New York School of Dental Medicine, Buffalo, NY, USA

Key points

- Periodontal diseases are common globally with up to 90% affected by the reversible form, gingivitis, and around half of adults having moderate or severe periodontitis with irreversible loss of soft and hard tissues surrounding the teeth.
- Periodontitis increases glycemic level and contributes to hyperglycemia/prediabetes, incident type 2 diabetes, diabetes complications, and possibly gestational diabetes.
- Periodontal treatment improves the glycated hemoglobin level by about 0.4 percentage points, similar to adding a second oral diabetes drug to metformin.
- Periodontitis and uncontrolled/poorly controlled diabetes adversely affect each other.
- Individuals with diabetes are more prone to periodontitis, and extent and severity of periodontitis increases with increasing levels of hyperglycemia. Therefore, periodontitis was suggested as the “sixth complication of diabetes” 20 years ago.
- People with diabetes have lost more teeth than those without diabetes.
- People with diabetes who obtain dental cleanings incur lower costs for diabetes-related medical care, including inpatient, outpatient, laboratory, and pharmacy costs.
- To manage diabetes, it would be beneficial to patients and society for healthcare providers to collaborate in patient-centered teams that include dental professionals.

Introduction

Periodontal disease

The vast majority of the world’s population are affected by periodontal disease with estimates of up to about 90% when the reversible form, gingivitis, is included [1]. The Global Burden of Diseases, Injuries, and Risk Factors Study (GBD) is the only comprehensive effort to estimate summary measures of global population health by cause, providing comparable estimates of the burden of 291 diseases and injuries in 1990, 2005, and 2010 [2]. Oral conditions affect 3.9 billion people, and untreated caries in permanent teeth is the most prevalent condition evaluated by GBD, with a global prevalence of 35% for all ages combined. During the 20 years from 1990 to 2010, the number of years lived with disability (YLD) increased for periodontal disease by 57.3%, whereas YLD for edentulism decreased by 12.4% [3]. This is in accord with people living longer and also keeping their teeth into older age. In 2010, severe periodontitis came in at sixth place of the most common diseases in the world, after dental caries of permanent teeth (1st), tension-type headache (2nd), migraine (3rd), fungal skin diseases (4th), and other skin and subcutaneous diseases (5th). The estimated 2010 global prevalence of the most severe category of periodontitis was just over 10% (10.79%), affecting men (10.89%) and women (10.68%) equally [2,3].

This estimate agrees with the analysis of the most recent national, representative data collected in the United States (US) during the 2009 and 2010 waves of the National Health and Nutrition Examination Surveys (NHANES). It estimates that about half the population aged 30 years or older have some form of periodontitis, either mild (~10%), moderate (~30%), or severe (~10%) [4]. Furthermore, 45% the US population 45–64 years old and 60% of 65–74 year olds have moderate or severe periodontitis with variation between race/ethnic groups [5].

For example, among Puerto Ricans 70 years and older, 83% have moderate (50%) or severe (24%) periodontitis [6]. Even in developed countries with a high standard of living, periodontal disease is prevalent. For instance, in a random sample of 1115 Danish 35–44 and 65–74 year olds, only 7.7% and 2.4%, respectively, had healthy periodontal conditions [7].

In susceptible individuals, periodontitis will progress with increasing loss of the soft and hard tissues surrounding the tooth. Unfortunately, this process may occur without any noticeable signs and symptoms. In the absence of any periodontal
treatment, loosening of the teeth may occur and the ultimate consequence would be their loss.

**Diabetes**

Worldwide, an estimated 382 million (8.3%) people currently have diabetes mellitus, with almost half of them (175 million) undiagnosed. Another 316 million with impaired glucose tolerance (IGT) are at risk. By 2035, the global population with diabetes is projected to rise by 55% to 592 million with another 471 having IGT, totaling 1 billion people. These are the ominous 2013 estimates by the International Diabetes Federation [8]. About 80% of persons with diabetes live in low- and middle-income countries, and in any country, the socially disadvantaged are most prone to suffer from this disease. Rapid lifestyle changes are bringing obesity and T2DM to historically high levels in countries in the Middle East, Western Pacific, sub-Saharan Africa, and South-East Asia. The highest numbers of people with diabetes live in China (98.4 million), India (65.1 million), and the USA (24.4 million) [8], with the latter estimate updated to 29.1 million, including 8.1 million undiagnosed in the US alone [9].

Global health expenditures due to diabetes for people 20–79 years of age are estimated for 2013 at USD 548 billion and for 2035 at USD 627 billion [8]. According to The World Health Organization (WHO), more than 80% of diabetes deaths occur in low- and middle-income countries, and diabetes will be the seventh leading cause of death in 2030 [10]. Based on health examination surveys and epidemiologic studies conducted from 1980 through 2008 among 2.7 million participants 25 years and older in 199 countries and territories, glycemia and diabetes were found to be rising globally, driven by both population growth and aging and by increasing age-specific prevalences [11].

Periodontal disease and diabetes are common, chronic, multifactorial diseases that share many risk factors [12]. The prevalence of both increases sharply with age and they often occur in the same individuals. Longevity is increasing in most countries and there is a tendency for people to keep their teeth longer than in the past. Hence, the need for treatment of both diseases, especially in older age groups, is also increasing and will likely continue to do so. About two-thirds of the US population 65 years and older have periodontitis [4] and more than one-quarter (25.9%) of this age group have diabetes [9].

Only recently are we beginning to unravel the mechanisms underlying both diseases and their mutual relationship and how they interact. Currently, it is likely that a major mechanism in this relationship is the systemic inflammatory response. Understanding the association between periodontal infection and diabetes has increased dramatically in recent years, as illustrated by the surge in the number of citations in PubMed over the last decade (Figure 68.1). Interprofessional management of the individuals with these chronic diseases should be our goal. Through creating a patient-centered medical home for each individual, proper management could benefit all involved.

**Periodontal diseases: assessment**

In the clinical situation, many factors will play a role in the decision made mutually by the dental professional and the patient regarding any treatment and management of gingivitis and periodontitis. Unfortunately, periodontitis is usually symptom free in the early and intermediate stages. Most people in the world have gingivitis at one or more sites in the mouth, but only some will experience bleeding during tooth brushing, flossing, or mastication of hard food items. Not until considerable breakdown of bone manifests itself as migration (turning, tilting, moving) or loosening of the tooth (mobility)—or suppuration causing observable flow or bad taste—will the patient become aware of the disease. At that time, it is often too late to salvage the tooth. In addition to observable signs and symptoms, patient-specific factors will be taken into account when a treatment plan is developed. Such factors include age, education, comprehension, ability and willingness to practice effective oral hygiene, frequency of check-up visits, personal values placed on oral health and teeth, as well as financial issues of payment and insurance coverage.

Figure 68.2 illustrates clinical measurement of various attributes of the periodontal tissues [13]. Color (reddening) or spontaneous bleeding in the gingival tissues, tooth mobility, or suppuration are often recorded. A common measure is bleeding upon probing (BOP), either tooth, surface, or site based, and a score can be allocated to each tooth, each tooth surface, or at each site that was probed for periodontal pocket depth. Complicating matters is that periodontitis, that is, irreversible breakdown of periodontal soft and hard tissues, occurs in very limited areas. The standard is probing at six different sites around each tooth. Periodontitis is site specific, which means that measurements just millimeters apart can yield substantially different results. Therefore, it is crucial to measure several sites for each surface [14]. Because the breakdown does not occur evenly or predictably on teeth, tooth surfaces, or sites on each surface, it is difficult to create valid and reliable statistical weights that can account for such variation when only a few sites in the entire dentition are measured for time and convenience reasons.

Assessment of periodontal disease may also be based on radiographic examination of the entire dentition at once (panorama radiograph), of segments of the dentition (bitewing radiographs), or of individual teeth (periapical radiograph that also shows the tip of the root and its adjacent jawbone). Due to the chronic nature of bone breakdown, it is not possible to determine whether the disease is active, based on radiographic measures at one time point, which could represent past bone loss. Assessment of antibodies as well as DNA identification of certain bacteria most often associated with periodontitis in saliva or serum may also be used, but these tests have not been validated.

Unfortunately, there are no detailed diagnosis codes in most countries for the two most common oral diseases, namely dental caries and periodontal diseases. Instead, codes for services provided are subsequently interpreted as an underlying
Figure 68.1 Number of PubMed citations on periodontal disease and diabetes by publication year: about half of the 2278 papers were published the last 8 years.

diagnosis for a condition in need of treatment having been present prior to the treatment rendered. Moreover, most countries do not have centralized databases for oral health. Consequently, it is very difficult to track or estimate the prevalence and trends in dental diseases in most populations.

To complicate matters, the resources needed to conduct population-based studies, or even large-scale clinical periodontal examinations, are time consuming and expensive. Properly trained and calibrated dental professional examiners and persons to record the measurements or codes called out (or alternatively voice-recognition devices) are required. Additionally, a portable examination chair or other seating, good lighting, sterile (disposable) instruments, and various disposables are needed. It would be unethical to withhold information from the examinees regarding any significant findings for which effective treatment is known to exist, so referrals would often be made. In order for the researchers to conduct meaningful statistical analyses later, responses to a written questionnaire or to an oral interview are needed to determine age, sex, socioeconomic level, medical history, and use of medications (for instance, those that influence inflammation such as statins and omega-3). In addition, responses
Figure 68.2 Clinical assessment of periodontal disease: terminology. CAL, clinical attachment level (“loss”) = PD minus CEJ; CEJ, cement–enamel junction; FMG, free marginal gingiva; PD, periodontal probing depth (= PPD); recession = CEJ minus FMG. Source: Clerehugh et al. 2009 [13]. Reproduced with permission of John Wiley & Sons.

regarding professional and home dental care, medical care, diet and exercise, and habits such as smoking and snuff use are required. Moreover, a brief physical examination is necessary to obtain data on weight, height, waist circumference, blood pressure, and glucose or glycated hemoglobin.

Such examination can easily last an hour or more per person, depending on the number of teeth and the extent of measures of various aspects of periodontal diseases. Further complicating matters is the fact that there are currently no globally accepted, standard case definitions for surveillance of periodontal diseases. Consequently, studies are not directly comparable; many research groups determine what will be considered periodontal disease in their specific study. The following example will illustrate this important, basic point, which is often not
sufficiently emphasized. In 1296 pregnant women, Manau and colleagues collected data on dental plaque, BOP, periodontal probing depth, and clinical periodontal attachment loss [15]. To this data set, they applied 14 different, previously published case definitions for periodontal disease, using over 50 continuous variables. The result was that between 3.2% and 70.8% (mean 35.8±(21.6)%; median 29.7%) of these women had periodontal disease, depending on which of the 14 case definitions was used.

However, recent case definitions developed for surveillance of periodontal disease by a joint workgroup of Centers for Disease Control and Prevention (CDC) and the American Academy for Periodontology (AAP) are used globally and are increasingly regarded as the standards [16,17]. Application of identical case definitions in different groups, populations, and countries would greatly enhance the comparability of findings on prevalence. In addition, this would enable estimation of trends within each population studied.

The lack of uniformly used definitions and attention to exactly how periodontal disease is defined must be kept in mind when interpreting results of studies related to periodontitis, including the evidence for the bidirectional effects of periodontitis and diabetes.

Self-reported periodontal disease

Of great interest and importance to healthcare providers, policy makers, administrators, and epidemiologists alike is the search for more economical modes of assessment of both individual and population-based risk for periodontal disease. There is increasing evidence for the validity and reliability of estimation of risk for periodontitis using self-report [18]; that is, using responses collected by simply asking individuals some questions that are shown to associate with the actual clinical status. Based on a pool of items formerly validated in population studies, the joint CDC/AAP workgroup has developed, cognitively tested, and applied in the field in Australia and the USA a set of eight such questions in English and Spanish [19,20]. In the 2009 and 2010 waves of NHANES, two important developments occurred in people 30 years and older for the first time ever: (1) all 28 teeth (excluding third molars) were clinically periodontally examined at six sites each (vs. formerly half-mouth, two or three sites per tooth), and (2) the eight CDC/AAP self-report questions regarding periodontitis were included in the questionnaire. The responses were obtained during in-home interviews with greater than 95% response rates and later validated against full-mouth clinically assessed periodontitis in 3743 adults. The self-reported measures performed well in predicting periodontitis in US adults. Furthermore, the questions and their answers achieved good results in predicting both the absence of any periodontitis and the presence of various levels of periodontitis [20]. The eight questions concern gum health and gum treatment history, loose teeth, bone loss around teeth, tooth not looking right, and use of dental floss and mouthwash. Table 68.1 displays these questions. It should be noted that the items are not yet ready for direct use in clinical practice to predict the risk for periodontitis in individuals, because such an index would require allocating weights to the item responses and demographic information, as well as field-testing and validation. However, the questions may be useful in assessing population-based levels of disease once validated in populations other than the US or Australian populations already tested [20,21].

**Periodontal diseases and inflammation: localized and generalized**

The two most common periodontal diseases are gingivitis, a reversible infection of the gingival tissues that surround the tooth, and periodontitis that is a multifactorial permanent breakdown of the supporting soft and hard tissues around the tooth. Both diseases are initiated by the microbial biofilm that accumulates on the tooth. Since the tooth consists of the only non-shedding tissue in humans, the biofilm can persist and organize itself over an extended period of time, provided external forces, such as tooth brushing or flossing do not disturb the dental plaque.

Whether the responding inflammation progresses to periodontitis depends on a multitude of factors mostly determined by the host [22]. Hence, periodontitis develops most often in susceptible individuals, and is not yet able to be predicted. Therefore, periodontitis can be regarded as an indication that the patient may have systemic conditions and diseases that deserve scrutiny and possibly referral for medical assessment and treatment [23].

The microbial biofilm will cause infection and local inflammation in the gingival tissues. The accompanying swelling will provide deepened spaces, so-called periodontal pockets, around the tooth, which will enhance the overgrowth of the biofilm. With progressing infection, the local inflammatory response escalates into cascading systemic responses during which the inflammatory mediators are disseminated throughout the body. Some of the commensal anaerobic bacteria will easily penetrate the inflamed tissue and enter the systemic circulation, in which they may be carried by phagocytic cells. Some of these bacteria and their endotoxins (lipopolysaccharides) further exacerbate the systemic anti-inflammatory processes. Wherever these bacteria lodge themselves, they may be able to invade and multiply, colonizing tissues and organs at body sites remote from the oral cavity. The subsequent local effect further adds to the systemic cascades of inflammatory responses.

This inflammatory response contributes to insulin resistance with elevation of blood glucose levels. Periodontal infection contributes to the cumulative burden of infection and its subsequent inflammatory responses. This is the main mechanism thought to underpin the effect of periodontal disease and systemic diseases, including diabetes and cardiovascular diseases. It has been shown that periodontitis can also contribute to development of prediabetes and T2DM—even possibly gestational diabetes.
Periodontal disease and diabetes mellitus

Table 68.1 CDC/AAP* questions for self-reported periodontal disease used in NHANES 2009 and 2010

<table>
<thead>
<tr>
<th>NHANES Item Name</th>
<th>Item Verbatim in English and Spanish (italicized) (Response Categories)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OHQB010</td>
<td>Do you think you might have gum disease? ¿Piensa usted que tal vez sufra de la enfermedad de las encías? (Yes, No, Refused, Don’t Know)</td>
</tr>
<tr>
<td>OHQB020</td>
<td>Overall, how would you rate the health of your teeth and gums? En general, ¿cómo diría que es el estado de salud de sus dientes y encías? (Excellent, Very good, Good, Fair, Poor, Refused, Don’t Know)</td>
</tr>
<tr>
<td>OHQB030</td>
<td>Have you ever had treatment for gum disease such as scaling and root planing, sometimes called “deep cleaning?” ¿Alguna vez ha tenido usted tratamiento de las encías tipo raspado o alisado de las raíces, que a veces se conoce como “limpieza profunda”? (Yes, No, Refused, Don’t Know)</td>
</tr>
<tr>
<td>OHQB040</td>
<td>Have you ever had any teeth become loose on their own, without an injury? ¿Alguna vez se le ha aflojado algún diente por sí solo sin haber tenido una lesión? (Yes, No, Refused, Don’t Know)</td>
</tr>
<tr>
<td>OHQB050</td>
<td>Have you ever been told by a dental professional that you lost bone around your teeth? ¿Alguna vez le ha dicho un profesional de la salud dental que usted ha perdido hueso alrededor de los dientes? (Yes, No, Refused, Don’t Know)</td>
</tr>
<tr>
<td>OHQB060</td>
<td>During the past three months, have you noticed a tooth that doesn’t look right? En los últimos tres meses, ¿ha notado usted un diente que no parece verse bien? (Yes, No, Refused, Don’t Know)</td>
</tr>
<tr>
<td>OHQB070</td>
<td>Aside from brushing your teeth with a toothbrush, in the last seven days, how many times did you use dental floss or any other device to clean between your teeth? Aparte del cepillado de sus dientes, ¿cuántas veces ha usado la seda/hilo dental o algún otro medio o utensilio para limpiarse entre los dientes en los últimos siete días? (___: Number of days, 77=Refused)</td>
</tr>
<tr>
<td>OHQB080</td>
<td>Aside from brushing your teeth with a toothbrush, in the last seven days, how many times did you use mouthwash or other dental rinse product that you use to treat dental disease or dental problems? Aparte del cepillado de sus dientes, ¿cuántas veces ha usado un enjuague bucal u otro producto líquido para el tratamiento de enfermedades o problemas dentales en los últimos siete días? (___: Number of days, 77=Refused)</td>
</tr>
</tbody>
</table>


Conceptual model of pathways linking periodontitis to diabetes

Figure 68.3 illustrates conceptually the main pathways in which periodontal inflammation contributes to diabetes and glycemic control, including hyperglycemia in people without diabetes or with prediabetes [24,25]. Periodontitis represents a chronic infection that contributes to the local and systemic cumulative bacterial burden. The inflammatory responses by the host to this periodontal infection are likewise chronic and manifest themselves as low-grade inflammation that maintains a role as a source for proinflammatory cytokines, such as interleukin-6 (IL-6), tumor necrosis factor-alpha (TNF-α), and interleukin-1-beta (IL-1β). These and other inflammatory mediators play important roles in the pathogenesis of insulin resistance by disrupting insulin-signaling leading to reduced uptake of glucose by the cells. Periodontal infection also leads to secretion by the liver of acute-phase reactants, such as C-reactive protein (CRP), fibrinogen, and plasminogen activator inhibitor-1 (PAI-1). The reactants in turn contribute to insulin resistance leading to hyperglycemia and prediabetes, and ultimately manifest T2DM. Epidemiologic studies have identified these acute-phase reactants as risk indicators and risk factors for the prevalence and incidence of both atherosclerotic cardiovascular disease and diabetes. Being a component of the metabolic syndrome, insulin resistance is recognized as a risk factor in the pathogenesis of diabetes and its complications, as well as cardiovascular disease. Adipokines and various other factors that contribute to insulin resistance are clearly important but are omitted from this model for simplicity of presentation.

Periodontal disease and diabetes mellitus: a two-way relationship

Periodontitis and poorly controlled diabetes negatively influence each other in a two-way causal relationship. Twenty years ago, periodontitis was suggested as the sixth complication of diabetes [26], and ample evidence now exists to show that periodontitis is indeed a nonspecific infectious complication of poorly- or un-controlled diabetes. The first systematic review
of the evidence for the opposite effect—namely, the effect of periodontal disease on hyperglycemia, prediabetes, diabetes, and diabetes complications—was published in 2013 [27,28].

Until recently, we believed that people with diabetes in general were more prone to periodontal disease, based on the view that individuals with diabetes have a compromised immune system and hence are less able to fight infections. However, only poorly or uncontrolled diabetes causes periodontal problems. A large number of studies have addressed this issue (Figure 68.1) and there is now a better understanding of the association between periodontal disease and diabetes. Examples of studies of different designs will be described to provide a summary of the depth and breadth of the body of evidence for the reciprocal relationship between periodontitis and diabetes.

Oral-systemic evidence reviewed by experts from Europe and the USA

Reports from a joint workshop comprising experts from the European Federation of Periodontology (EFP) and the American Academy of Periodontology (AAP) assessed the current scientific evidence for associations between periodontal disease and diabetes, cardiovascular disease, adverse pregnancy outcomes, and other systemic diseases. These reports may be freely accessed at: http://onlinelibrary.wiley.com/doi/10.1111/jcpe.2013.40.issue-s14/issuetoc?dmmsmid=73654&dmmspid=17129767&dmmsuid=1955451 (last accessed July 15, 2014). Four of these reports concern diabetes, namely three peer-reviewed reviews on underlying pathologic mechanisms [29,30], effect of periodontal treatment on glycemic control [31,32], and effect of periodontal disease on diabetes [27,28]. The fourth is a consensus statement summarizing the evidence regarding periodontal disease and diabetes [33,34]. It also contains recommendations regarding oral healthcare in people with diabetes, which will be described at the end of this chapter. Additionally, the results of the systematic review of epidemiologic observational evidence for the effect of periodontal disease on diabetes [27,28] are summarized in a publicly accessible webcast at: http://www.scivee.tv/node/58235 (last accessed July 15, 2014).

Longitudinal studies from which causality may be deducted are rare. Evidence regarding causal relationships between periodontal disease and diabetes stems mostly from longitudinal clinical, serologic, and radiologic studies among the Pima Indians in Southwestern USA. These Native Americans have an extraordinary high prevalence of T2DM. The studies showed in this population with poorly controlled diabetes and poorly managed periodontal disease that T2DM was associated with highly increased risks for periodontal disease. However, only recently has it become evident that it is not having a diagnosis of diabetes mellitus per se that causes increased prevalence and severity of periodontitis, it is the hyperglycemia in un- or poorly controlled diabetes that most likely is the offender [12].

The fact that individuals with well-controlled diabetes are not at greatly elevated risk for periodontal disease is demonstrated in several epidemiologic observational and interventional studies. For example, among 4343 NHANES 1988–1994 participants aged 45 years and older of whom 502 had diabetes (fasting plasma glucose $\geq 126$ mg dL$^{-1}$), severe periodontitis was found in 14.3% of those with poorly controlled diabetes ($\text{HbA1c} > 9\%$) and in 8.8% of those with “better” controlled ($\text{HbA1c} \leq 9\%$) diabetes, compared to 4.1% among people without diabetes [35]. After controlling for potential confounders, individuals with diabetes who had poor glycemic control had 2.90 times, and those with better control 1.56 times, higher risk for periodontitis than individuals without diabetes. Only the former obtained statistical significance. These findings support the notion that individuals with well-controlled diabetes have about the same prevalence of periodontitis as otherwise similar individuals without diabetes, when keeping in mind that the threshold for “good” diabetes control currently would likely be lower than the 9% HbA1c used in this report [35]. Furthermore, persons with well-controlled diabetes usually respond well periodontally to routine, nonsurgical periodontal treatment.

Effect of diabetes on periodontal disease

In the studies among Pima Indians, the age of onset, severity, and prevalence of periodontal disease are all adversely affected by T2DM [36,37]. In fact, individuals with diabetes were up to four times more likely to develop periodontal disease than their counterparts without diabetes [38]. These studies confirmed and extended a 20-year-old landmark study in a Finnish population that showed that poorly controlled diabetes increases the risk for progressive periodontal bone loss more than diabetes under good glycemic control does [39].
A large-scale epidemiologic survey in Germany assessed multiple potential risk factors and supports the associations between both T1 and T2DM and periodontitis [40].

People with diabetes have more periodontal disease, according to a 2009 systematic review and meta-analysis of 49 cross-sectional studies by Chavarry and colleagues [41]. Of these studies, 17 concerned type 1; 26 type 2; and 6 both T1 and T2DM. In another review of cross-sectional studies worldwide that included a control or comparison group without diabetes, the prevalence of periodontal disease was higher in the diabetes groups in the vast majority of the respective studies [24].

A dose–response relationship was demonstrated in a study of both T1DM and T2DM and periodontal disease in Columbia [42]. The higher the level of hyperglycemia, the more severe was the periodontitis. In a study of T1DM, more children and adolescents with diabetes had periodontal disease as compared to those with no diabetes OR = 2.24 (1.02–4.93) [43].

About one third of the US adult population with diabetes have moderate or severe periodontal disease with young adults being about twice as likely — and those 45 years and older almost three times as likely — as their peers without diabetes to have severe periodontitis [44].

The Chavarry 2009 systematic review and meta-analysis mentioned earlier included eight longitudinal studies, namely four observational and four interventional studies among participants with T1DM and T2DM [41]. The authors conclude that T2DM is an independent risk factor for periodontitis.

In the large, longitudinal population based study of 2558 participants between 20 and 81 years of age in Western Pomerania in northern Germany, the progression of periodontitis was monitored over a period of 5 years [45]. Importantly, diabetes was demonstrated to be a significant risk factor for incident clinical attachment loss as well as increased periodontal probing depth. The following other risk factors were identified: male sex, age 65 and older, lower education level, and being single or divorced. Hence, the authors urge offering periodontal prophylaxes or maintenance sessions especially targeted to those groups who experienced the greatest progression of periodontitis.

Type 1 diabetes

Children who are between 6 and 14 years old have a mixed dentition, that is, both primary teeth (baby teeth) and permanent teeth in various stages of eruption are present in the oral cavity. Studying 270 such children with diabetes and comparing them to 320 children without diabetes, it was shown that the former exhibited accelerated tooth eruption in the late mixed dentition period at ages 10–14 [46].

In an earlier study including 142 patients 11 to 18 years of age with T1DM, it was found that 9.8% suffered from periodontitis, as compared to 1.7% of a comparable control group of 116 subjects with no diabetes [47]. Another study was conducted among 18–50-year-old patients in India with T1DM for at least 4 years and compared them to healthy controls matched for age, physical build, and oral hygiene [48]. The mean HbA1c level was 8.8% in the diabetes group. It was demonstrated that all periodontal disease measures (gingival bleeding, periodontal probing depth, and clinical attachment loss) were more severe in T1DM patients, despite currently having similar oral hygiene levels. Periodontal destruction assessed as mean clinical attachment level was also greatly worse in subjects with longer disease course (4.1 vs. 2.8 mm), poorer metabolic control (4.4 vs. 2.2 mm), and the diabetes complications retinopathy, neuropathy, and nephropathy (4.2 vs. 2.3 mm).

Lalla and colleagues demonstrated that periodontal breakdown can begin very early in life in children and adolescents with diabetes [49,50], and gingival bleeding is more common in children with diabetes compared to those without diabetes [51]. Furthermore, this accelerated periodontal destruction in children is related to metabolic control. Hence, it is likely important to attain and maintain good glycemic control to minimize periodontal disease, just as is the case for other diabetes complications [52].

Effect of diabetes intervention on periodontitis in type 2 diabetes

One recent report investigated the reverse effect, that is, the effect of interventions to improve glycemic control on the periodontal disease in persons with T2DM. This is a novel concept to be studied. In Tokyo, 35 persons between 40 and 75 years and HbA1c of at least 7.4% and with periodontitis underwent therapy to improve glycemic control — without periodontal treatment. Glycated hemoglobin (HbA1c), high-sensitivity C-reactive protein (hsCRP), and periodontal measures were assessed at baseline, and 2 and 6 months after the therapy to improve glycemic control. The authors concluded that effective glycemic control improves periodontal bleeding on probing in people with T2DM [53].

Effect of periodontal disease on diabetes

Evidence from cross-sectional studies clearly shows that individuals with periodontitis, especially if severe and of great extent, have much higher risks for elevated blood glucose or glycated hemoglobin than those with a healthy periodontium. However, such evidence is able to show only associations, but not temporality. In early 2013, the first systematic review of epidemiologic, nonexperimental, observational evidence for the effect of periodontal disease on hyperglycemia, prediabetes, and diabetes was published [27,28]. The systematic review only considered studies in which the temporal sequence of the effect of periodontitis on glycemic control could be assessed and hence included mostly longitudinal studies. However, studies on diabetes complications were also included because such complications are not known to cause periodontal disease and therefore, the temporal sequence was given. Seventeen studies were analyzed and
they support significant, adverse effects of periodontal disease on glycemic control, diabetes complications, and development of type 2 (and possibly gestational) diabetes. Importantly, periodontal disease seems to lead to elevated glucose levels even in healthy individuals free of diabetes. Major limitations were: small studies, variability of exposure and outcome parameters, and limited generalizability of the results [27,28].

The notion that periodontal infection would adversely affect blood glucose levels is biologically plausible, as this is a normal part of the inflammatory response to infection as it occurs anywhere in the body. Periodontal infection likely contributes to the cumulative burden of infection and its subsequent systemic inflammatory response. Moreover, periodontal infection leads to systemic dissemination of members of the oral microbiome. Several of these microorganisms and their endotoxins induce strong systemic inflammatory responses in susceptible individuals, which in turn contribute to insulin resistance. It was demonstrated in a longitudinal study of 3280 participants free of diabetes and obesity at baseline, that individuals who later developed diabetes indeed had higher levels of bacterial DNA in their blood than those who did not develop diabetes [54]. The source and role of these bacteria are not known, but they may originate from the oral, gut, or skin microflora.

To further strengthen the evidence for the effect of periodontal infection on development and control of diabetes and its complications, observational cohort studies are needed in larger populations for longer periods. In addition, intervention studies in which baseline HbA1c is elevated (uncontrolled glycemic levels) should demonstrate that periodontal treatment is effective in reducing systemic inflammation to finally establish that periodontal treatment that reduces the periodontal infection will also reduce HbA1c levels.

### Effect of periodontal treatment on diabetes

Because periodontal infection contributes to the cumulative infection burden and its associated inflammatory responses, it should logically follow that elimination—or prompting a major decrease of the infectious burden—should lead to a decrease in blood glucose levels.

Increasingly solid evidence is emerging for routine, nonsurgical periodontal treatment leading to significant decrease in glycated hemoglobin at 3 months post treatment, and Table 68.2 displays all such meta-analyses conducted. The resulting decrease of between 0.4% and 0.8% HbA1c actual difference is not only statistically significant, but also clinically relevant in diabetes management. The magnitude of this improvement upon periodontal treatment is comparable to that seen by adding a second oral antidiabetic medication to metformin [33,34]. It is noteworthy that these meta-analyses include a 2010 review by the Cochrane Collaboration [55] and that they all come to largely the same conclusion.

The effects of periodontal therapy on HbA1c levels in patients with poorly controlled diabetes are illustrated in Table 68.3, which displays detailed (unpublished) data from the three

### Table 68.2 Effect of nonsurgical periodontal treatment on glycemic control in people with type 2 diabetes: meta-analyses published as of December 11, 2013

<table>
<thead>
<tr>
<th>Year &amp; Ref.</th>
<th>#Studies</th>
<th>#RCT</th>
<th>Pooled # Subjects</th>
<th>HbA1cChange</th>
<th>95% CI</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Janket et al. 2005</td>
<td>4</td>
<td>1</td>
<td>264</td>
<td>−0.66%</td>
<td>−2.2; 0.9</td>
<td>ns</td>
</tr>
<tr>
<td>Darre et al. 2008</td>
<td>9</td>
<td>9</td>
<td>485</td>
<td>−0.46%</td>
<td>0.11; 0.82</td>
<td>0.01</td>
</tr>
<tr>
<td>Teeuw et al. 2010</td>
<td>5</td>
<td>3</td>
<td>180</td>
<td>−0.40%</td>
<td>−0.77; −0.04</td>
<td>0.03</td>
</tr>
<tr>
<td>Simpson et al. 2010 (Cochrane Review)</td>
<td>3</td>
<td>3</td>
<td>244</td>
<td>−0.40%</td>
<td>−0.78; −0.01</td>
<td>0.04</td>
</tr>
<tr>
<td>Sgolastra et al. 2013</td>
<td>5</td>
<td>5</td>
<td>315</td>
<td>−0.65%</td>
<td>−0.43; −0.88</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>Engebretson &amp; Kocher 2013</td>
<td>9</td>
<td>9</td>
<td>775</td>
<td>−0.36%</td>
<td>−0.54; −0.19</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Liew et al. 2013</td>
<td>6</td>
<td>6</td>
<td>422</td>
<td>−0.41%</td>
<td>−0.73; −0.09</td>
<td>0.013</td>
</tr>
</tbody>
</table>

| RCT, randomized controlled trial; HbA1c, A1c glycosylated hemoglobin; CI, confidence interval; 1Remainder of the studies (non-RCT), clinical controlled trials; | 1Standardized mean difference (SMD); ns, non-significant. |

### References for Table 68.2

Periodontal disease and diabetes mellitus

Table 68.3  Change in glycated hemoglobin (HbA1c) level 3 months after nonsurgical periodontal treatment by pretreatment HbA1c level (n = 57)*

<table>
<thead>
<tr>
<th>Pretreatment HbA1c</th>
<th>n</th>
<th>HbA1c change 3 months post treatment (±SD)</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;7.0%</td>
<td>4</td>
<td>0.93 (±1.74)</td>
<td>ns</td>
</tr>
<tr>
<td>7.0–7.9%</td>
<td>7</td>
<td>−0.17 (±1.03)</td>
<td>ns</td>
</tr>
<tr>
<td>8.0–8.9%</td>
<td>7</td>
<td>−0.49 (±0.81)</td>
<td>p &lt; 0.05</td>
</tr>
<tr>
<td>9.0–9.9%</td>
<td>6</td>
<td>−0.98 (±1.79)</td>
<td>ns</td>
</tr>
<tr>
<td>10.0–10.9%</td>
<td>8</td>
<td>−1.18 (±1.04)</td>
<td>p &lt; 0.05</td>
</tr>
<tr>
<td>11.0–11.9%</td>
<td>8</td>
<td>−0.45 (±0.62)</td>
<td>p &lt; 0.05</td>
</tr>
<tr>
<td>12.0–12.9%</td>
<td>10</td>
<td>−1.17 (±0.91)</td>
<td>p &lt; 0.05</td>
</tr>
<tr>
<td>≥13.0%</td>
<td>7</td>
<td>0.81 (±0.93)</td>
<td>p &lt; 0.05</td>
</tr>
</tbody>
</table>

*Additional data not included in the report published by Grossi et al. [56]; HbA1c, glycated hemoglobin; n, number of subjects; ns, not statistically significant at p < 0.05; p, probability; SD, standard deviation.


Table 68.3 shows statistically significant reductions at 3 months post intervention (Table 68.3). It is therefore reasonable to expect a greater reduction in glycated hemoglobin after periodontal therapy in patients whose initial HbA1c level is high.

These results are clinically significant as well. The CDC estimates that “in general, every percentage point drop in A1c blood test results (e.g. from 8.0% to 7.0%) can reduce the risk of microvascular complications (eye, kidney, and nerve diseases) by 40%” [44].

Effect of periodontal disease on diabetes complications

The relationship between periodontal disease and other complications of diabetes has been addressed by several studies. In individuals with T1DM and T2DM, those without teeth or with periodontal disease, especially severe disease, have higher risk for diabetes-related complications than those without or with little periodontal disease [23,27,28,57,58]. A dose–response effect seems to exist between severity of periodontal disease and risk for complications. For instance, it was found that individuals with diabetes suffered from periodontitis that was more severe if they also experienced other complications, such as retinopathy [59]. Assessment of microvascular hemorrhaging at two distinct anatomical sites was conducted only in the 1988–1994 NHANES [60]. From this study it was found that individuals with bleeding in one or more of five gingival sites had a 57% increased odds ratio for also having retinal hemorrhaging. Modeling showed that 51% of this association was explained by degree of hyperglycemia, assessed by HbA1c levels. A Japanese study also reported the severity of periodontal disease was significantly correlated with the severity of diabetic retinopathy (p = 0.0012). Moreover, the risk of proliferative diabetic retinopathy was significantly higher (2.8 times; p = 0.036) in the presence of periodontal disease [61].

Of the seven studies identified in our systematic review [27,28], one reported on T1DM [62] and five on T2DM [63–67], whereas one did not specify diabetes type [61].

Longitudinal studies among Pima Indians 35 years and older demonstrate that periodontal disease may predict increased overall mortality [65] as well as the risk of overt nephropathy and end-stage renal disease [66]. Periodontal disease significantly predicted deaths from ischemic heart disease (p_trend = 0.04) and diabetic nephropathy (p_trend < 0.01). Those with severe periodontal disease had 3.2 times the risk (95% CI 1.1–9.3) of cardiorenal mortality (ischemic heart disease and diabetic nephropathy combined) compared to those with no, mild, or moderate periodontal disease combined [65]. Similarly, during up to 22 years of follow-up, the incidence of macroalbuminuria was 2.0, 2.1, and 2.6 times higher in individuals who had moderate periodontitis, severe periodontitis, or were edentulous, respectively, compared to those with no or mild periodontitis (p = 0.01) [66]. Comparing the same groups, incidence of end-stage renal disease was 2.3, 3.5, and 4.9 times higher (p = 0.02). Hence, moderate and severe periodontal disease, as well as edentulousness, significantly predict in a dose-dependent manner both macroalbuminuria and end-stage renal disease among Pima Indians.

Diabetes complications reported to be related to periodontal disease include:

- cardiovascular disease
  - coronary artery disease
  - ischemic heart disease
  - carotid intima media thickening
  - calcification of atherosclerotic plaque
- hypertension
- cerebrovascular events (stroke)
- kidney disease
  - macroalbuminuria
  - proteinuria
- end-stage renal disease
- death due to cardio-renal disease
- neuropathic foot ulcer
- retinopathy.

Periodontal disease and diabetes-related medical care costs

Medical care costs related to periodontal care

Diabetes care requires immense resources as mentioned earlier. For instance, about USD 8500 is spent annually per diabetes...
patient for diabetes care in the US [8]. The most recent estimated US diabetes costs among individuals with diagnosed diabetes in 2012 was 2.3 times higher per patient than for those without diabetes [9]. The total US diabetes costs are USD 245 billion with USD 176 billion spent on direct medical costs and USD 69 billion in indirect costs (disability, work loss, premature death) [9].

Periodontal disease is relatively easy to prevent or manage, as the diseased tissues are easily accessible and treatment occurs at outpatient dental offices or clinics. In addition, inexpensive home-care measures are known to be effective in management or prevention of recurrence. Therefore, it would be of potentially wide-ranging interest to explore whether periodontal care is associated with reduced expenditure for diabetes-related medical services, actually reflecting the potential decreases in healthcare costs. The area of exploring dental and medical insurance claims is a novel avenue to obtain a glimpse of such information. Brief examples of analyses of insurance claims in US studies with different medical care outcomes follow.

Hospitalizations and physician office visits
Among 91,242 insureds with T2DM and both medical and dental insurance, the annual mean number of hospitalizations over a 4-year period was reduced by 39.4%, in the insureds who received periodontal care when compared to insureds who did not receive such care [68]. Provision of monthly oral health procedures by licensed dental hygienists to residents of nursing homes [69] reduces the hospitalization rate by over half, namely 8.8% to 11.7%, compared to the national rate of over 25% [70]. Several studies demonstrate that medical care costs for patients with diabetes are lower for those who obtain preventive or therapeutic periodontal care than for people with diabetes who do not receive periodontal care. For instance, in people with T2DM, costs for medical care were significantly reduced by USD 2840 (from USD 7056 to USD 4216) per subject per year, a 40.2% reduction, for individuals who also received periodontal care compared to those who did not [68]. Another study among insureds with diabetes found that those who received actual periodontal treatment (a proxy for the diagnosis of manifest periodontitis, not only gingivitis) incurred significantly higher per member per month medical costs compared to insureds who only received gingivitis treatment, dental maintenance services, other dental services, or no dental services [71]. However, it was not possible to control for potential confounders such as smoking and socioeconomic factors as the insurance database lacked this information. Furthermore, among 2674 insureds aged 18–64 with diabetes who had simultaneous medical, pharmaceutical, and dental insurance coverage for at least 1 year during a 6-year period, those who received the anti-inflammatory, nonsurgical periodontal procedures: debridement, scaling and root planing (“deep cleaning”), periodontal maintenance therapy, or prophylaxis had significantly lower medical care costs compared to the diabetes group not receiving these dental services for diabetes-related care; diabetes complications care; and total medical care [72].

Periodontal care and glycemic control
Among 5103 dually (medical and dental) insureds 40–70 years old with diabetes, the mean glycated hemoglobin level was 7.66% [73]. The 38% who received some type of periodontal treatment had an average glycated hemoglobin level that was 0.08 percentage points higher than those who did not (p = 0.02), likely reflecting the inflammatory effect of periodontal infection. Patients with greater periodontal treatment intensity had significantly lower glycated hemoglobin levels than those receiving treatment of lower intensity. In those with multiple surgeries for treatment of periodontal disease, glycated hemoglobin levels were reduced by 0.36 percentage points (p = 0.002) compared to those with no or one periodontal surgery procedure. These findings support the findings from the clinical trials of effects of periodontal treatment on glycemic control (Tables 68.2 and 68.3).

Guidelines for oral health in people with diabetes
Hyperglycemia and periodontal infection adversely affect each other: Poorly controlled diabetes is associated with the incidence, progression, and severity of periodontitis. Similarly, periodontitis decreases glycemic control, increases the risk of diabetes-associated complications, promotes incident diabetes (onset of new diabetes)/impaired fasting glucose, and possibly promotes gestational diabetes. Periodontal treatment—including routine, nonsurgical procedures that can be delivered in general dental practices and clinics—can lead to better glycemic control as well as reduce periodontal inflammation and promote periodontal health.

It is estimated that among the 29.1 million Americans aged 20 years and older who have diabetes, 21.0 million have diagnosed and 8.1 million undiagnosed diabetes [9]. Additionally, 86 million have prediabetes [9]. Consequently, over 115 million—a staggering proportion exceeding one third of the entire adult US population—are affected by diabetes or by a high risk for diabetes. Therefore, even a small improvement in hyperglycemia by attaining and maintaining good oral health could have great beneficial effects on diabetes control, complications, costs, and human suffering. Consequently, prevention or treatment of periodontal infections should become an integral part of the management of diabetes.

International Diabetes Federation (IDF)
For use by people with diabetes and their care providers, the International Diabetes Federation has published guidelines for oral health specifically in people with diabetes [74]. This 11-page booklet named IDF Guideline on Oral Health for People with Diabetes is freely available at: http://www.idf.org/webdata/docs/OralHealth_EN_RTP.pdf.

American Diabetes Association (ADA)
Gradually, but slowly, oral health is becoming part of the standards for diabetes care around the world. For instance, up to
and including 2013, the annually updated standards of medical care in diabetes by the American Diabetes Association (ADA) include only referral to a dentist for comprehensive periodontal examination without any further comments or explanations [75]. For the first time, a brief section with two citations on periodontal disease appeared in the 2014 annual January supplement to Diabetes Care, on p. 749 in the updated “Standards of Medical Care in Diabetes — 2014”, which may be freely viewed at: http://care.diabetesjournals.org/content/37/Supplement_1/S14.full.pdf+. This paragraph is included in the section “VII. Assessment of Common Comorbid Conditions.” It is a positive step by the ADA towards recognizing the importance of oral health in diabetes management.

European Federation of Periodontology (EFP) and the American Academy of Periodontology (AAP)

A consensus report regarding the relationships between periodontal disease and diabetes including suggestions regarding oral care for patients with diabetes and for physicians and dentists who treat patients with diabetes, or patients at risk for diabetes, resulted from the joint European–USA conference of experts [33,34]. These recommendations are displayed in Figure 68.4. The guiding principles can also be publicly viewed at: http://onlinelibrary.wiley.com/doi/10.1111/jcpe.12077/pdf.

Co-management of patients with diabetes by dental and medical care providers

Patient-centered collaboration between healthcare professionals to prevent, control, and manage both periodontitis and diabetes is a reasonable goal since good glycemic control and good oral hygiene in people with diabetes will decrease the risk for development of periodontal disease and tooth loss. Conversely, attaining and sustaining periodontal health will prevent increased blood glucose levels due to periodontal infection. Despite the importance of assessing the oral cavity for (especially periodontal) infection as an integral part of the comprehensive medical care of diabetes patients, such attention seems to still be generally lacking in the medical community. However, the dental profession is increasingly concerned with the importance of good periodontal health in patients with diabetes as well as considering the possible presence of undiagnosed diabetes in these patients.

Focus on oral health in medical practice: why medical health professionals should pay attention to the health of the mouth

The health of and care for the oral cavity has become increasingly separated from the rest of the body. Although the concept of the strong oral-systemic connection is not new, it was often lost in our modern specialization. However, recent attention has focused on the mutually important effects between the mouth and the body and novel techniques are now available to explain some of the mechanisms underlying such relationships. Both periodontal disease and diabetes are chronic diseases that increase with age, and it is now known that they mutually influence each other and share common risk factors such as poor diet, obesity, and smoking. Consequently, it becomes imperative to make every effort to ensure that good oral health is attained and maintained. This is increasingly important as more people keep their teeth into old age and there is a tendency to be less willing to part with their teeth.

Many studies show that people with diabetes feel that their oral health negatively affects their quality of life [76–78]. A Brazilian study among persons with diabetes even demonstrated that oral health affected those who had mild to moderately severe periodontal disease more negatively than those with healthy periodontal tissues or only gingivitis [79].

There is very limited interprofessional education among the healthcare professions in most countries regarding the oral cavity and the ways in which systemic diseases and conditions manifest themselves in the oral tissues—often prior to exhibiting signs elsewhere in the body.

Several reports from the Institute of Medicine (IOM) and other professional associations and foundations emphasize the need for better integration of healthcare in the future and the various communities of health professionals are beginning to realize the potential benefits of such an integration [80]. For instance, the concept of a medical home with patient-centered interprofessional collaboration to manage patients is slowly gaining ground [81,82].

Focus on diabetes in the dental office: should dental health professionals screen for diabetes?

About 70% of the US population visit a dentist within a year [83], and it is likely that this is the case in many other developed countries. Therefore, it is reasonable to propose the dental visit may offer a largely untapped opportunity to screen for undiagnosed diabetes and prediabetes, which could lead to referral of patients for further medical evaluation and care. Especially beneficial would be the early detection of elevated risk of hyperglycemia, prediabetes, since early lifestyle intervention may prevent or delay the onset of diabetes. It is estimated that among the approximately 25% of people in the US who do not receive any general outpatient healthcare visits, over one quarter did visit a dental office or clinic. Therefore, dental practices can serve as alternate sites of opportunity to identify general health concerns [84].

The American Diabetes Association recommends that testing to detect T2DM and assess risk for future diabetes in
A Suggested Guidelines for Physicians and Other Medical Health Professions for Use in Diabetes Practice

Oral health of my patients

Because of the increased risk for developing periodontitis in patients with diabetes, the following recommendations are made:

- Patients with diabetes should be told that periodontal disease risk is increased by diabetes. They should also be told that if they suffer from periodontal disease, their glycaemic control may be more difficult, and they are at higher risk for diabetic complications such as cardiovascular and kidney disease.
- As part of their initial examination, patients with type 1, type 2 and gestational diabetes (GDM) should receive a thorough oral examination, which includes a comprehensive periodontal examination.
- For all newly diagnosed type 1 and type 2 diabetes patients, subsequent periodontal examinations should occur (as directed by the dental professionals) as part of their ongoing management of diabetes. Even if no periodontitis is diagnosed initially, annual periodontal review is recommended.
- Diabetes patients presenting with any overt signs and symptoms of periodontitis, including:
  - loose teeth not associated with trauma
  - spacing or spreading of the teeth
  - gingival abscesses
  - gingival suppuration
- require prompt periodontal evaluation.
- Patients with diabetes who have extensive tooth loss should be encouraged to pursue dental rehabilitation to restore adequate mastication for proper nutrition.
- Oral health education should be provided to all patients with diabetes.
- For children and adolescents diagnosed with diabetes, annual oral screening is recommended from the age of 6–7 years by referral to a dental professional.
- Patients with diabetes should be advised that other oral conditions such as dry mouth and burning mouth may occur, and if so, they should seek advice from their dental practitioner. Also patients with diabetes are at increased risk of oral fungal infections and experience poorer wound healing than those who do not have diabetes.

B Suggested guidelines for use in dental practice

My patient with diabetes

- Patients with diabetes should be told that they are at increased risk for periodontitis. They should also be told that if they suffer from periodontal disease, their glycaemic control may be more difficult, and they are at higher risk for other complications such as cardiovascular and kidney disease.
- Patients presenting with a diagnosis of type 1, type 2 or gestational diabetes should receive a thorough oral examination, which includes a comprehensive periodontal examination.
- If periodontitis is diagnosed, it should be properly managed. If no periodontitis is diagnosed initially, patients with diabetes should be placed on a preventive care regime and monitored regularly for periodontal changes.
- Patients with diabetes presenting with any acute oral/periodontal infections require prompt oral/periodontal care.
- Patients with diabetes who have extensive tooth loss should be encouraged to pursue dental rehabilitation to restore adequate mastication for proper nutrition.
- Oral health education should be provided to all patients with diabetes.
- Patients with diabetes should also be evaluated for other potential oral complications, including dry mouth, burning mouth and candida infections.
- For children and adolescents diagnosed with diabetes, an annual oral screening for early signs of periodontal involvement is recommended starting at the age of 6 years.
- Patients who present without a diabetes diagnosis, but with obvious risk factors for type 2 diabetes and signs of periodontitis should be informed about their risk for having diabetes, assessed using a chair-side HbA1C test, and/or referred to a physician for appropriate diagnostic testing and follow-up care.

C Recommendations for patients with diabetes at the physician's practice/office

Why should I have my gums checked?

If your physician has told you that you have diabetes, you should make an appointment with a dentist to have your mouth and gums checked. This is because people with diabetes have a higher chance of getting gum disease. Gum disease can lead to tooth loss and may make your diabetes harder to control.

You may have gum disease if you have ever noticed:

- Red, bleeding or swollen gums;
- Pus from the gums;
- Foul taste;
- Longer looking teeth;
- Loose teeth;
- Increasing spaces between your teeth;
- Calculus (tartar) on your teeth.

If you have noticed any of these problems, it is important to see a dentist as soon as possible. Gum disease may be present and get worse with no apparent symptoms to you, so even if you do not think you have gum disease now, you should still get regular dental check-ups as part of managing your diabetes. Your dentist will be able to pick up early signs of gum disease.

You also need to clean your teeth and gums very carefully at home. If you have diabetes, you may also suffer from dry mouth, burning mouth, yeast infections of the mouth or poor healing of mouth wounds. It is important to keep your mouth and your whole body as healthy as possible with regular dental and medical care.

D Recommendations for patients at the dental surgery/office who have diabetes or are found to be at risk for diabetes

You have diabetes or you have been told by your dentist you are at risk for diabetes

People with diabetes have a higher chance of getting gum disease. If you have been told by your dentist that you have gum disease, you should follow up with necessary treatment as advised. This may require several appointments. Like diabetes, gum disease is a chronic condition and requires lifelong maintenance.

You also need to clean your teeth and gums very carefully at home.

If left untreated, gum disease may also make your diabetes harder to control. Gum disease may be present and get worse with no apparent symptoms to you, so if your dentist told you that you do not have gum disease now, you should still get regular dental check-ups as part of managing your diabetes. Your dentist will be able to pick up early signs of gum disease.

You may have gum disease if you ever notice:

- Red, bleeding or swollen gums;
- Pus from the gums;
- Foul taste;
- Longer looking teeth;
- Loose teeth;
- Increasing spaces between your teeth;
- Calculus (tartar) on your teeth.

People with diabetes may also suffer from dry mouth, burning mouth, yeast infections of the mouth or poor healing of mouth wounds. If you do not have diabetes, but your dentist identified some risk factors for diabetes including signs of gum disease, it is important to get a medical check-up as advised.

Your medical doctor can order blood tests to see if you have diabetes and not know it, and can provide proper advice and care based on the results. Make an appointment to see your medical doctor as soon as possible. And remember to inform your dentist about the outcome of your visit to the medical doctor.

It is important to keep your mouth and your whole body as healthy as possible with regular dental and medical care.
asymptomatic people should be considered in adults of any age who are overweight or obese (BMI ≥ 25 kg m⁻²) or have one or more additional risk factors for diabetes. In those without these risk factors, testing should begin at age 45 years [75].

Several studies suggest a patient with periodontal disease is at risk for diabetes [85–89]. Therefore, presence of periodontal disease can be used as an additional criterion for screening for diabetes in the dental office [23]. Lalla and colleagues demonstrated that having at least 26% of teeth with deep pockets or missing at least four teeth correctly identified about three-quarters (72–75%) of prediabetes or diabetes cases [87].

Several surveys indicate that dental professionals are interested in and willing to engage in evaluation, screening, and monitoring systemic conditions and diseases, as well as to collaborate more closely with other healthcare professionals. For example, findings from a study assessing dentists’ attitudes, willingness, and perceived barriers regarding chair-side medical screening in the dental office show that of the 1945 responding dentists, about three-quarters (76.6%) regarded it important to conduct screening for diabetes, with similar results for other systemic conditions such as hypertension (85.8%), cardiovascular disease (76.8%), hepatitis (71.5%), and human immunodeficiency virus infection (68.8%) [90]. The vast majority were willing to refer patients for consultation with physicians (96.4%), collect oral fluids for salivary diagnostics (87.7%), and conduct medical screenings that yield immediate results (83.4%). About half would collect blood via finger stick (55.9%). Respondents were significantly more willing to collect saliva than height and weight measurements or blood via finger stick. Insurance reimbursement turned out to be significantly less important than time, cost, liability, or patients’ willingness. Confirmation of the latter was demonstrated among elder minorities during community-based oral health screening sessions that integrated screening for diabetes and hypertension into its oral health activities at senior centers [81].

A recent study screening dental patients for diabetes reported that about 40% of patients attending a dental office visit were found to have elevated HbA1c (≥5.7%) when referred for medical diagnosis. Only 21% from dental offices compared to 85% from Comprehensive Care Clinics sought the medical diagnosis. Only 21% from dental offices compared to 85% from Comprehensive Care Clinics sought the medical diagnosis from a physician after referral. This study identified possible barriers, that is, dental patients’ hesitancy to follow up with medical referrals may be due to costs, denial, or some other factors. These barriers did not seem to exist in the Community Care Clinic. Factors that facilitate and encourage follow-up with diabetes referral in similar studies may help us understand these barriers and develop strategies to overcome them [91].

For dental professionals to engage in identification and management of prediabetes and diabetes and to collaborate closely with other health professionals requires a paradigm shift, as well as proper education [92]. Attitudes of dental care providers, medical professionals, and patients and the individuals engaged in their health will need to shift towards regarding all professionals who engage in attaining and maintaining the patient’s good health as members of an interprofessional team. One approach is to engage dental health professionals in managing risk factors common to periodontal disease and diabetes. These common risk factors include smoking and diet, especially refined sugar consumption [93]. We all need to see ourselves as partners who assume important roles in the patient-centered healthcare team whose collaboration is necessary for obtaining the goal of the best health possible for each patient.

**Future healthcare collaboration for diabetes prevention and lifelong management**

In recent history, health professions’ education and healthcare practices have developed and functioned separately, with little recognition of the fact that they are all inextricably linked. However, rapid redesign of healthcare delivery, stimulated in the US in part by the Affordable Care Act, is occurring alongside, but independently of, educational reform of the health professions. For instance, health professionals participating in a conference in January 2013 envisioned a healthcare system in which learners and practitioners across the professions are working collaboratively with patients, families, and communities, and with each other [94]. The Healthy People Curriculum Task Force is a consortium of eight health professions education associations. It has developed The Education for Health Framework and connected the framework with new and revised educational objectives of Healthy People 2020. Interprofessional prevention education, in which health professionals learn and practice together, is seen by the Task Force as a key method for implementation [95].

Team care is likely to become an important component of future healthcare systems designed to provide comprehensive lifetime prevention and management of chronic diseases such as diabetes, in which dental professionals may play an important role. The “team care” guide named *Redesigning the Health Care Team: Diabetes Prevention and Lifelong Management* is a 48-page document created by the US National Diabetes Education Program [96]. These programs are designed for the US healthcare system and may not be ideal for other countries, but in general, integrated interprofessional healthcare is important and likely to benefit patients suffering from diabetes and periodontitis throughout the world. The guide may be accessed at: http://ndep.nih.gov/media/NDEP37_RedesignTeamCare_4c_508.pdf.

Physicians’ provision of more comprehensive and effective oral health advisory and routine communication with dentists is part of the expanded physician role consistent with the evolving patient-centered medical–dental home or “chronic care model” of the future [97,98]. Similarly, dental professionals should be vigilant regarding their patients’ general health and refer as necessary for medical evaluation.

The time has come for all healthcare professionals to collaborate and co-manage their mutual patients with diabetes—or at risk for developing diabetes—for potential human and financial benefits for the patients and societies globally.
References


33. Chapple IL, Genco R; Working group 2 of joint EFP/AAP workshop: Diabetes and periodontal diseases: consensus report of the


70 Lobo AA, Braner L: The hospital admission rate for residents serviced by PrevMED programs in 2012; 4pp. 2013. This report may be obtained from: Ellsworth Harris, President, PrevMED - Onsite Health Care Services; PrevMED Corporate Offices, 210 Pier One Road STE 201, Stevensville, Maryland 21666. Tel.: +1.410.604.3350; e-mail: info@monthlyservice.com; http://www.prevmed.org/contact-us/ (accessed July 27, 2014).


Complications of diabetes: macrovascular
CHAPTER 69

Epidemiology of macrovascular disease and hypertension in diabetes mellitus

Jaakko Tuomilehto1,2,3, Daiva Rastenyte5, Qing Qiao1,2, Noël C. Barengo2 and Karl Matz2
1National Institute for Health and Welfare, Helsinki, Finland
2Department of Public Health, University of Helsinki, Helsinki, Finland
3Danube University Krems, Krems, Austria
4King Abdulaziz University, Jeddah, Saudi Arabia
5Lithuanian University of Health Sciences, Kaunas, Lithuania

Introduction

Diabetic patients suffer from cardiovascular disease (CVD) and the majority of both type 1 diabetes mellitus (T1DM) and type 2 diabetes mellitus (T2DM) patients die from CVD. Epidemiologic studies of CVD in diabetes are important for the estimation of the magnitude and the nature of the problem, for the identification of the risk factors for CVD in diabetic subjects, and for the development of preventive measures. It has been proposed that diabetes and macrovascular disease may both have common genetic or environmental antecedents. This “common soil” hypothesis for diabetes and macrovascular diseases has received more support from the numerous epidemiologic studies during the last few years [1]. The major forms of CVD, that is, coronary heart disease (CHD), stroke and peripheral vascular disease (PVD), share both an atherosclerotic background and several common risk factors.

Hypertension

Prevalence and incidence

It has been known for decades that the prevalence of hypertension in diabetic individuals appears to be approximately twice as high as in nondiabetic people in the same population [2–4]. It is more frequent in men than in women before the fifth decade and more frequent in women thereafter [5]. The prevalence of coexisting diabetes and hypertension is higher in US Blacks compared with Whites; both conditions are more common among the lower socioeconomic groups [6–8].

The prevalence of hypertension among young T1DM patients is dramatically increased. Data from the Joslin clinic showed that 15.5% of males with diabetes aged 18–24 years had hypertension compared with only 1.7% in the general population—excess risk of 13.8% [9]. A much lower prevalence of hypertension was found in females in the same age group: 2.4% versus 0.8%, respectively. Others have also reported a significant increase in blood pressure (BP) in T1DM males but not in T1DM females [10]. In the general population BP levels are higher in men than in women until the age of 45–50 years, and this may also be reflected in young diabetic subjects. Nevertheless, other studies suggest that significant increases in BP in T1DM patients are seen in both sexes [11].

Epidemiologic evidence is unequivocal and uniform, showing that patients with T2DM have a higher prevalence of hypertension than do nondiabetic subjects at any age, in both sexes and regardless of BP levels in the background population [12]. The problem for the interpretation of the available epidemiologic data is that in many studies control for degree of obesity was not done or remained unclear. The increased prevalence of hypertension in the diabetic population (especially T2DM) may to some extent be due to improved screening since diabetic patients use more health services, and due to the lower thresholds for treatment of high BP in diabetic patients [13].

The definition for hypertension has changed over the years, and also varies to some extent between countries. The most recent criteria of the US Eight Joint National Committee (JNC8) were issued in 2014 [14]. In general, the JNC8 and World Health Organization/International Society of Hypertension (WHO/ISH) define hypertension as an office BP exceeding 140
and/or 90 mmHg (Korotkoff I–V). The JNC8 states that there is strong evidence to support treating hypertensive persons aged 60 years or older to a BP goal of less than 150/90 mmHg and hypertensive persons 30 through 59 years of age to a diastolic goal of less than 90 mmHg; however, there is insufficient evidence in hypertensive persons younger than 60 years for a systolic goal, or in those younger than 30 years for a diastolic goal, so the panel recommends a BP of less than 140/90 mmHg for those groups based on expert opinion. The same thresholds and goals are recommended for hypertensive adults with diabetes or nondiabetic chronic kidney disease (CKD) as for the general hypertensive population younger than 60 years [14].

There are a few studies that have shown that in T2DM the increase in BP is, at last in part, independent of body weight [12,15,16]. While T1DM patients are usually normotensive until overt renal disease develops, the prevalence of hypertension increases markedly in T2DM before the proteinuric state [17,18]. About one fourth of T2DM patients are already hypertensive at the time of diagnosis of diabetes [3,17,19].

**Proteinuria/Microalbuminuria and BP**

Microalbuminuria is in most cases the first sign of the diabetic renal disease in both T1DM and T2DM patients. The prevalence of microalbuminuria among diabetic patients is 15–20% [20–22] and it is higher in T1DM than T2DM patients. On the other hand, more or less severe signs of diabetic renal disease may be found in as many as 41% of both T1DM and T2DM patients once the duration of diabetes is long enough [22]. It is possible that high BP in diabetic patients with either microalbuminuria or nephropathy is genetically determined, but the susceptibility genes have not been identified yet. There is little question that BP is higher, even within the normal range, in T1DM patients with microalbuminuria compared with normoalbuminuric patients [23]. However, whether BP rises before or after the development of microalbuminuria has been debated. Mathiesen et al. [24] found no difference in BP between those who did and those who did not develop persistent microalbuminuria until microalbuminuria had been present at least 3 years, indicating that BP elevation is secondary to the initiation of diabetic nephropathy. Raal et al. [25] found no difference in BP at baseline between individuals who remained normoalbuminuric and those who later progressed to diabetic renal disease. Nevertheless, the development of microalbuminuria was accompanied by a parallel rise in diastolic BP, indicating that BP elevation and the initial stage of diabetic renal disease are concurrent events. Kimball et al. [23] also suggested that the alterations in cardiovascular and renal function may occur in parallel in adolescents with T1DM. However, some studies suggested that BP may rise before the detection of microalbuminuria and may even serve as a predictor of its development [26–28]. Results from the EURODIAB T1DM Complications Study [29] also suggest that raised albumin excretion is a response to, and not necessarily a cause of, mildly raised BP.

The pathogenetic relationship between elevated BP and renal disease in T2DM is more difficult to establish than in T1DM. Often the time of onset of T2DM cannot be precisely defined, and some 40–50% of T2DM patients receive antihypertensive treatment at the time of diagnosis [30]. It has been shown that elevated BP is an independent predictor of T2DM and glucose intolerance [4]. However, BP starts to increase steeply when persistent proteinuria develops in T2DM patients [31], and evidence exists that BP may start to rise before the development of microalbuminuria not only in T2DM patients but also in T1DM patients [27,31]. It is possible that in many hypertensive patients with T2DM, diabetes itself is not causally related to elevated BP, since even in nondiabetic populations of the same age hypertension is very common.

**Hyperinsulinemia/insulin resistance and BP**

Numerous conditions associated with insulin resistance and/or hyperinsulinemia are frequently accompanied by hypertension. This conclusion applies not only to obesity [32,33], impaired glucose tolerance (IGT) [34], or T2DM [35], but also to T1DM [36] when insulin sensitivity is impaired [37]. It has been suggested that hyperinsulinemia, which occurs with insulin resistance, increases BP primarily through an increase of sodium retention, stimulation of sympathetic nervous activity, or both [38]. However, epidemiologic and clinical data linking hyperinsulinemia to hypertension are inconsistent. There is support for the insulin–BP association in some Caucasian populations [34,39], case-control studies [40–42], or cross-sectional studies [43,44], but a number of studies in ethnic groups including Europids [45–47], Micronesians, Polynesians, and Melanesians [48], Pima Indians [49,50], African Americans [49], Asian Indian, Creole, and Chinese Mauritians [51] do not support the hypothesis. Interestingly, these latter populations have substantially higher circulating insulin concentrations than do, for instance, Europid populations, and yet, their BP levels are not higher or may be even somewhat lower. In addition, insulin resistance has not always been demonstrated in patients with essential hypertension. Patients with severe hyperinsulinemia due to an islet cell tumor, for example, display neither hypertension nor hyperglycemia [52]. Likewise, the supraphysiologic dose of insulin administered to patients with T1DM is not always associated with hypertension [53]. One study indicated a correlation between insulin and BP only in subjects with a positive family history of hypertension or diabetes and suggested a possible genetic influence on insulin resistance [54]. A role of a familial component in insulin resistance syndrome is also suggested by studies showing that young adult normotensive offspring of hypertensive parents are more insulin resistant than offspring of normotensive parents [55]. Doria et al. [56] suggested that insulin resistance occurs
only in those hypertensive subjects with high erythrocyte sodium–lithium countertransport.

Inconsistency in association between insulin and hypertension can be in part explained by the hemodynamic effects of insulin. There is a strong evidence that acute physiologic increments in plasma insulin concentrations (euglycemic clamp) stimulate the sympathetic noradrenergic activity, which is offset by its vasodilator action on skeletal muscle blood flow [57]. The acute vasodilatory effect of insulin may be through alteration of intracellular calcium handling. Saito et al. [58] found that insulin attenuated angiotensin II stimulated increases in cytosolic free calcium by decreasing the pool of inositol triphosphate (IP$_3$) releasable calcium, an effect blocked by inhibition of nitric oxide generation. More recent studies show an important effect of insulin on the vascular production of endothelin 1 [59].

The opposing hemodynamic effects of insulin help explain why acute physiologic hyperinsulinemia does not elevate BP in normal human beings [60]. Baron et al. [61] hypothesize that insulin resistance and hyperinsulinemia are markers of the hypertensive process in the muscle and not contributors to elevated BP. A sympathetic overactivity could also lead to insulin resistance and a compensatory increase in plasma insulin levels [62,63]. The recent line of research related to insulin has been targeted to endothelial dysfunction [64]. Hyperinsulinemia (or insulin resistance) seems to abrogate endothelium-dependent vasodilation in large arteries. One of the proposed mechanisms includes oxidative stress caused by hyperinsulinemia.

Attenuation in the balance between insulin’s depressor and pressor effects could influence BP regulation. Insulin-mediated stimulation of skeletal muscle blood flow is reduced in obese subjects [65] and T2DM patients [66]. However, several clinical studies have failed to show any effect of acute physiologic hyperinsulinemia on BP in obese, insulin-resistant hypertensive patients [67–69]. Thus it is probable that hyperinsulinemia is not directly causing hypertension, but may serve as an indicator of a common antecedent (either genetic or environmental) of T2DM and hypertension. In studies of Mexican Americans a number of factors including typical CVD risk factors such as high total cholesterol, BP, fasting glucose, body mass index (BMI), triglycerides, and low high-density lipoprotein (HDL) cholesterol predicted the development of T2DM, but their effect was virtually removed after adjusting for fasting insulin [70]. This study strongly suggested that prediabetic individuals who are insulin resistant are at a higher risk of developing CVD than are insulin-sensitive prediabetic individuals, pointing out the importance of hyperinsulinemia in targeting prevention [71].

**Hyperglycemia and BP**

Hyperglycemia is the distinguishing feature of diabetes. The joint data from Finnish and Dutch cohorts of the Seven Countries Study with 30 years of follow-up showed that in men with T2DM or IGT, BP levels and the prevalence of hypertension were higher than in normoglycemic men. Interestingly, these differences in BP had already been seen 20–30 years earlier. An elevation in BP preceded abnormal glucose tolerance but not hyperinsulinemia. BMI levels of men with hyperinsulinemia had been shown to be higher 20–30 years earlier, but BP levels had not. Thus, it was proposed that glucose intolerance is a stronger correlate of hypertension than is hyperinsulinemia [72].

The hypothesis of oxidative stress through which high glucose levels may play a role in cardiovascular complications has been increasingly discussed [73,74]. The endothelium has been shown to release a substance that induces smooth muscle relaxation by increasing the production of cyclic guanosine monophosphate. This factor is called endothelium-derived relaxing factor, and nitric oxide is believed to be one of such factors [75]. Vessels of healthy people are continuously diluted by nitric oxide released from endothelial cells. Exposure to elevated concentrations of glucose in vitro causes selective impairment of endothelium-dependent relaxation [76]. Oxidative stress induced by hyperglycemia is implicated as a source of altered endothelially mediated relaxation in diabetes. It was shown in experimental studies that oxygen-derived free radicals inactivate endothelium-derived releasing factors [77] and selectively attenuate endothelium-dependent relaxation [78]. One possible source of oxygen-derived free radicals in diabetes is autoxidation of glucose [79]. Glucose can autoxidize, generating free radicals, hydrogen peroxide, and reactive ketoaldehydes. These last compounds may largely participate in the formation of glycated proteins, which are themselves a source of oxygen-derived free radicals [80,81]. The possibility that oxygen-derived free radicals play a role in the pathogenesis of vascular complications of diabetes is suggested by studies that have shown that antioxidants such as vitamin E, superoxide dismutase, catalase, glutathione, and ascorbic acid are all decreased in tissues and blood of diabetic patients [82,83]. Also elevated levels of products of oxygen-derived free radicals have been reported in diabetic patients [84,85]. In a study by Ceriello et al. [86], the plasma concentration of O$^-$ was elevated in patients with T1DM, but showed a trend to normalization after strict metabolic control. There was a strong correlation between plasma glucose concentrations and O$^-$ concentrations in both normal subjects and diabetic patients over a wide range of glucose concentrations. An acute antihypertensive effect of antioxidant agents, such as vitamin C, thiopronine, and glutathione, has been reported in hypertensive subjects, regardless of whether they are diabetic patients [86]. During the postprandial state, not only hyperglycemia occurs, but also hyperlipidemia. More recently, it was confirmed that both postprandial hypertriglyceridemia and hyperglycemia have independent and cumulative deleterious effect on endothelial function and suggested that oxidative stress is the common mediator through which they exert such effect [87].
Diurnal blood pressure variability

BP and heart rate are principally under the control of the autonomic nervous system. It has been proposed that loss of the normal day–night difference in BP may be a sign of early autonomic dysfunction in diabetic patients. In normotensive T1DM and T2DM patients during sleep there is loss of the nocturnal dip of BP seen in normal subjects, and diabetic patients also have increased BP variability [88–90]. These may be contributing factors to the development of hypertension and the accelerated cardiovascular damage seen in diabetes [88]. In addition, left ventricular mass is increased [91–94], asymptomatic cerebrovascular damage is more advanced [95], and history of stroke is more frequent [96] in nondiabetic hypertensive nondippers as compared with dippers. Thus, for any given level of daytime BP, a persistent pressure overload could speed the progression of hypertensive organ damage and thereby result in a higher incidence of cardiovascular events. Since cardiovascular risk in hypertension is somewhat lower in nondiabetic women than in men, the detrimental effect of persistently high BP levels over 24h would become particularly evident in diabetic women, thus raising their risk to the level occurring in men [97]. A study on determinants of ambulatory 24-h BP in T1DM patients has shown that the mean male–female difference in 24-h diastolic BP was 5.6 mmHg less in diabetic patients than in healthy individuals after adjusting for all other variables [98]. These findings are in concert with the reported loss in diabetic women of the relative female protection against cardiovascular and cerebrovascular diseases [99,100].

Atherosclerosis

Diabetic subjects show more widespread and more severe coronary atherosclerosis. Data supporting this opinion emerge equally from postmortem [101–103] and from clinical and epidemiologic studies [104]. The Five Town Study [105] found more extensive atherosclerosis in the coronary arteries and abdominal aorta of diabetic subjects than in the groups that represented “standardized average” atherosclerosis in the background population. In addition, there was a tendency for an increased severity of atherosclerosis when both diabetes and hypertension had been present in life. Another WHO study on the prevalence of large vessel disease among diabetic populations in 14 European centers suggested the following [106]:

1 The prevalence of large vessel disease varies considerably between different centers, the highest being in Eastern Europe. This corresponds to the prevalence of CVD in the general population.

2 The prevalence of all large vessel disease is at least the same in diabetic women as in diabetic men. The overrepresentation of women seems particularly great in centers where the overall prevalence of atherosclerotic diseases is high.

3 The female overrepresentation is due to CHD while the prevalence of PVD is higher in men than in women in most centers. The first statement is in agreement with the observations from the International Atherosclerosis Project [107].

At least two studies were designed to evaluate the severity of coronary artery disease specifically in juvenile-onset T1DM. One was an autopsy study, in which coronary pathology of nine young patients with T1DM was compared with a group of nine age-matched individuals without diabetes [107,108]. None of the individuals in either group had a clinical history of angina pectoris. Six of the nine diabetic patients had a >75% reduction in luminal diameter, while none of the nondiabetic subjects had this degree of narrowing. Another study examined clinical and angiographic findings in 32 patients with T1DM and 31 nondiabetic patients, matched for age and symptoms [109]. In comparison with nondiabetic individuals, patients with T1DM were significantly more likely to have severe (>70%) narrowings, to have them in all three major coronary arteries (47% vs. 6%), and to have them in distal segments. Severe narrowings of multiple vessels were significantly more common in men than in women and in individuals who also had hypercholesterolemia.

In the only population-based autopsy study, autopsies were evaluated in 83 diabetic and 159 nondiabetic Japanese American men from the Honolulu Heart Program [103]. Diabetic men had an excess of coronary atherosclerosis and of acute, healing, or fibrotic myocardial lesions; the latter but not the former association was independent of major CHD risk factors. This suggests that the atherosclerotic process is mediated by BP, cholesterol, and smoking, while in men with diabetes one or more modes of action besides atherosclerosis also seem to account for the excess of heart disease.

Results from other autopsy studies have shown that diabetic patients have a higher incidence of two- and three-vessel disease, and a lower incidence of single-vessel disease, compared with patients without diabetes. In a report by Natali et al., the incidence of multivessel disease was as high as 36% in individuals with diabetes compared with 17% in nondiabetic individuals (p < 0.0001) [110]. Retrospective analyses of patients undergoing elective percutaneous transluminal coronary angioplasty (PTCA) support these data. Stein and coworkers analyzed results from 1133 diabetic and 9300 nondiabetic patients undergoing PTCA and showed that the patients with diabetes had a higher rate of multivessel disease than did those without diabetes [111]. The Thrombolysis and Angioplasty in Myocardial Infarction Trial involved 148 diabetic patients and 923 nondiabetic patients [112]. Patients with diabetes had a significantly higher incidence of multivessel disease (66%) than did the patients without diabetes (46%, p = 0.0001). In addition, the number of diseased vessels per patient was greater among patients with diabetes compared with those without diabetes.

The effect of diabetes on cerebral and peripheral circulation has been studied extensively during the last decades. Chan et al. [113], using noninvasive methods, compared the prevalence of carotid artery stenoses (50% reduction or more) in 286 volunteer T2DM subjects and 135 controls of similar age (mean 61 and 60 years, respectively); 7% of the diabetic subjects and 1% of controls had severe stenoses. Pujia et al. [114] examined carotid arteries of 54 T2DM subjects and 54 sex- and
age-matched control subjects using echo-Doppler examination. The prevalence of carotid atherosclerosis was 46% in T2DM subjects and 18% in control subjects. Yamasaki et al. [115] assessed the carotid arteries in 105 patients with T1DM (age 4–25 years, duration of diabetes 0.5–17 years), 529 patients with T2DM (age 31–86 years, duration of diabetes 0.5–49 years), and 104 nondiabetic healthy subjects. The intimal plus medial thickness (IMT) values of the arterial wall for T1DM patients aged 10–25 years were significantly greater than those in age-matched nondiabetic subjects. T2DM patients showed IMT values equivalent to those in normal adults over 20 years old. A cross-sectional study on the relation of carotid artery stiffness indices with glucose intolerance and serum insulin concentrations found that persons with T2DM or borderline glucose intolerance had stiffer arteries than did their counterparts with normal glucose tolerance [116]. According to the results of the Insulin Resistance Atherosclerosis Study, diabetes was significantly associated with increased atherosclerosis in the internal carotid artery as determined by IMT [117]. Several studies reported a higher degree of early atherosclerosis in newly detected type 2 diabetic patients than in nondiabetic subjects [118,119]. This suggests that hyperglycemia together with clustering of risk factors, particularly dyslipidemia, may cause intimal–medial thickening already in the early phases of diabetes.

Kingsbury [120] studied 338 male nondiabetic subjects referred to a surgical clinic specializing in PVD. Arteriography was performed in each subject and the degree of irregularity classified into three grades of severity. Both fasting and 2-h blood glucose levels were related to the degree of irregularity, independent of age. The mean ages of men in the “moderate” and “extensive” categories of PVD were 59 and 61 years, respectively. Thus, nondiabetic men at increased risk of T2DM (because of relative glucose intolerance) seem to have more extensive PVD. In the Schwabing Study [121], more than 600 T1DM and T2DM patients were prospectively followed for low-extremity arterial disease, for example PVD by ultrasonic Doppler measurements. Both at the 5-year and 9-year follow-up, the incidence of mostly asymptomatic was consistently and significantly associated with baseline BP and the dyslipidemic complex of high-serum triglycerides/low-HDL cholesterol, but not with total cholesterol or weight. The association with HbA1, plasma C-peptide, and daily insulin dose (in insulin-treated T2DM) was weaker and at 9-year follow-up only significant in univariate analysis. Of diabetic patients dying from cardiovascular causes within 5-year observation, 67% had PVD (mostly asymptomatic) at baseline compared with 15% in those who survived.

Mozes et al. [122] compared the histomorphologic appearance of atherosclerosis in amputated legs of diabetic (10 type 2, 4 unclassified) and nondiabetic patients (n = 14). In diabetic patients occlusive disease in amputated legs was more severe in arteries above the ankle than in nondiabetic patients while no difference was found in this series in arteries of the ankle and foot. Also medial calcification in the arteries tended to be more prominent in diabetic patients than in nondiabetic patients [122]. Jude and collaborators [123] performed the first study that compared angiographic findings between diabetic and nondiabetic patients with PVD. This study showed that in a group of patients referred for angiography of the lower extremity, diabetic patients have worse angiographic findings, more amputations (41.4% vs. 11.5%), and higher mortality than do nondiabetic patients [123].

Many other studies confirm an association between diabetes and increased prevalence of PVD. Individuals with diabetes have 2–4-fold increase in the rates of PVD [124], more often have femoral bruits and absent pedal pulses [125], and have rates of abnormal ankle–brachial indices ranging from 11.9% to 16% [126,127]. The duration and severity of diabetes correlate with incidence and extent of PVD [123]. Patients with diabetes have a greater risk of intermittent claudication, gangrene (10-fold), and amputations (15–20-fold) [128,129]. In the Framingham cohort, the presence of diabetes increased the risk of claudication by 3.5-fold in men and 8.6-fold in women [130]. Overall, diabetes causes most nontraumatic lower extremity amputations in the United States [131]. The relative risk estimate for lower extremity amputation in patients with diabetes was 12.7 compared with that of nondiabetic patients in the Medicare population and as high as 23.5 for diabetic persons aged 65–74 years. Inconsistency exists regarding the question of whether PVD is more commonly symptomatic or asymptomatic among patients with diabetes [129,132].

**Coronary heart disease**

**Prevalence and incidence**

There are geographical differences in the prevalence of CHD among cohorts of diabetic patients, and in the incidence of CHD in the general population [133]. Although CHD is a relatively rare cause of death in the first 30 years of T1DM, it is much more common in later years, accounting for up to 30% of deaths in those with T1DM for more than 40 years. A study of 370 male employees with T1DM of a US company showed that death rates in diabetic subjects were more than double than in healthy controls during a 10-year period [134]. Mortality ratios were highest under the age 45, mainly because of increased risk of CHD. Researchers at the Joslin clinic reported that after 20–40 years duration of diabetes 33% of patients aged 45–59 years had either symptomatic or asymptomatic CHD [135]. This is likely to be an underestimate however, as only survivors were included in the study population. Four occupational/population-based studies from the United States found an increased risk of incident heart disease (fatal and nonfatal combined) among diabetic persons [104,136–138]. National survey data from the United States on adults with T2DM have shown a prevalence of fatal and nonfatal CHD events of around 4% for those aged 18–44 years and 20% for those older than 65 years [139]. These rates are 2–20 times higher than corresponding ones for
similar-aged people without T2DM, the rate-ratios being highest among younger patients with T2DM. Uusitupa et al. [140] have also found that age-standardized prevalence of acute myocardial infarction (AMI) among newly diagnosed T2DM men (17%) and women (18%) was significantly higher than among nondiabetic ones (10% and 4%, respectively). Also the incidence of first myocardial infarction (MI) was significantly higher among diabetic compared with nondiabetic persons (for men 27% vs. 19%; for women 12% vs. 2%) during the 10-year follow-up. Age-standardized cardiovascular mortality was substantially higher in diabetic men (15%) and women (17%) than in nondiabetic men (5%) and women (2%).

In one population-based study, the 7-year incidence of first MI or death for patients with diabetes was 20% but was only 3.5% for nondiabetic patients [141]. History of MI increased the rate of recurrent MI or cardiovascular death for both groups (18.8% in nondiabetic persons and 45% in those with diabetes). Thus, patients with diabetes but without previous MI carry the same level of risk for subsequent acute coronary events as nondiabetic patients with previous MI. This study and other findings were considered when the US National Cholesterol Education Program Adult Treatment Panel III proposed that more intensive interventions should be considered for diabetic than nondiabetic subjects to prevent CHD [142].

Many studies have shown a considerable 2–8-fold excess in CVD mortality in insulin-treated diabetic subjects compared with the general population [143–147]. Cardiovascular mortality has been estimated to be 2.5-fold higher in T2DM men and 2.9-fold higher in T2DM women compared with nondiabetic subjects. Those treated with oral hypoglycemic agents had higher rates of cardiovascular mortality than did those treated with diet alone, indicating that the severity of diabetes contributes to the risk of cardiovascular mortality [148,149].

Recent extensive evidence confirms the excess risk of cardiovascular mortality in diabetic patients. Data from the intensive blood-glucose treatment arm of the United Kingdom Prospective Diabetes Study (UKPDS) examined the reasons for death in patients with newly diagnosed type 2 diabetes over a 10-year follow-up period. The UKPDS showed that 8.4% of type 2 diabetic patients died of either fatal MI or sudden death, whereas mortality associated with other disease was relatively very low [150]. Data from an age- and sex-matched group in the Framingham Study show that for nondiabetic individuals the equivalent mortality rate over 10 years is only 4% [151].

Women in general have a lower unadjusted risk of CHD than do men. In many studies, rates of CVD in women with diabetes equal or exceed those in men [152–154], although not all studies have demonstrated this relationship [155,156]. More research is also needed to determine whether this is true for both T1DM and T2DM. The Wisconsin population-based study reported a similar risk ratio in both T2DM men (2.4) and women (2.2) versus nondiabetic persons. Nevertheless, in T1DM women the risk ratio (13.5) was significantly higher than that in T1DM men (9.1) [157]. The reason for the accelerated atherogenesis among diabetic women is not completely understood, but it may relate to hyperglycemia itself since this phenomenon is seen both in young-onset T1DM patients [158] and in older T2DM patients [104]. It may also relate, at least in part, to more severe lipid and lipoprotein abnormalities, particularly elevated levels of triglycerides and reduced levels of HDL, among diabetic women [159,160]. A recent study [161] suggests greater impairment of endothelial function associated with T2DM in women than in men. The role of hyperinsulinemia and insulin resistance is important in understanding the link between female gender and risk of CVD in diabetic patients. Insulin resistance is associated with lower estrogen and higher androgen levels. Premenopausal women with insulin resistance or diabetes do not benefit from the protective effects of menstruation experienced by women without these conditions [162]. In addition, the magnitude of increased risk of reinfarction and fatality rate following an acute MI among diabetic patients compared with nondiabetic patients was greater in women than in men [163]. Several studies have noted that CVD mortality in diabetic patients was related to age at the onset of diabetes or to duration of diabetes [145–147,164]. In the Diabetes Epidemiology Research International Study of young-onset T1DM patients, cardiovascular deaths were relatively few during a follow-up of 8–15 years, although significantly increased as compared with the rates in the background population [165].

Excess mortality, CVD mortality in particular, extends beyond the limits of diabetes to IGT. Studies in the United Kingdom, in Bedford [166] and in Whitehall, London [167], have all consistently shown increased mortality at the upper end of the blood glucose distribution before T2DM is manifest. This is in agreement with the findings of Uusitupa et al. [168] who showed that a considerable number of T2DM patients had already suffered an MI before the time of diagnosis of their diabetes. Similar findings were also observed in the Pacific Island populations where CHD is much less common than in Europid populations [169].

**Mortality and case-fatality in AMI**

There are wide differences in the age-adjusted incidence of fatal CHD in T2DM patients between populations: the highest rates were in Framingham [170] where they were 4% per year for men, but the rates were only 1% per year in the Chicago Heart [171] and MRFIT [172] studies. In Europe, the CHD incidence varied from 2.6% per year in Finland [173] and 2.2% in the Schwabing cohort [174], to 1 to 1.4% per year in the United Kingdom [175] and in another study in Finland [140] and 0.4% per year in the Paris Prospective Study in France [176]. Prospective studies from Japan have reported very low mortality from CHD in both T2DM and T1DM [165,177], corresponding to the general rates in Japan. Also in Pima Indians who have the highest prevalence of T2DM in the world, mortality from CHD
is not particularly high [178]. Nevertheless, the most recent information on total and CVD mortality among the Native Americans in the Strong Heart Study has revealed that currently American Indian populations have higher CVD mortality rates than in other ethnic groups in the United States [153].

A number of studies have demonstrated that diabetes mellitus negatively influences the survival after MI, including increased rates of reinfarction, congestive heart failure, and death [112,179–187]. The largest population-based study of the impact of diabetes on mortality after the first MI was the FINMONICA Study [185]. The 1-year case-fatality rate for first MI (including prehospitalization mortality) was 45% in diabetic men and 39% in diabetic women. These case-fatality rates were significantly higher than rates in nondiabetic male and female subjects (38 and 25%, respectively). Furthermore, diabetic patients were more likely than nondiabetic patients to die from sudden cardiac death before reaching hospital. In unstable angina pectoris or non-Q-wave in-hospital MI, the presence of diabetes increases the risk of in-hospital MI, complications of MI, and mortality [186]. In the OASIS Registry, a six-nation study of unstable angina and non-Q-wave MI, diabetes independently increased the risk of death by 57% [187]. In the SHOCK Trial of revascularization for MI complicated by cardiogenic shock, the relative risk of death for patients with diabetes was 1.36 compared with that of nondiabetic patients [188].

There also seems to be a differential gender-related effect regarding the type of diabetes on case-fatality in AMI. Data from the GISSI-2 Study have shown that in men, both T1DM and T2DM were associated with moderately higher mortality rate after MI compared with nondiabetic men [183]. In women, diabetes was a strong risk factor for death following an MI. In the Framingham Study [163], which included a much higher proportion of women than men taking insulin (27% vs. 15%), the relative risk of fatal CHD was higher for women than men. In men, the presence of diabetes may lead to an overall decreased survival rate independent from the type of diabetes, whereas in women, T1DM is uniquely associated with major alterations in cardiovascular function that have a strong impact on mortality after MI.

Reviewed by Jacoby and Nesto [189], the reasons for the increased case-fatality in AMI in diabetic patients still remain debated. Some of the studies suggested that a higher prevalence of other risk factors (hypertension, hyperlipidemia, extensive coronary artery disease, and advanced age) or higher degree of left ventricular damage after infarction may explain the higher mortality rate in diabetic patients. Conversely, other studies [190] indicate that the detrimental effect of diabetes on survival is most evident in patients whose baseline risk factor profile would indicate a relatively low risk of death after MI, thus suggesting that pathophysiologic derangements accompanying diabetes that apparently do not affect the clinical status of the patients may play a subtle but critical role as they reduce survival after MI. A study from the community-based AMI register in Finland demonstrated that there were no differences in cardiac enzyme or ECG findings between diabetic and nondiabetic patients who had their first AMI [184]. Thus, there was no evidence that the size of AMI is larger in diabetic than nondiabetic subjects. It is known that AMI is often associated with an acute hyperglycemia. This may be particularly aggravated in diabetic AMI patients leading to various metabolic and cardiac problems during the first 1–2 weeks after AMI. The DIGAMI Study suggests benefit from glucose lowering post-AMI [191], and so this problem may be reversible. The current evidence has also revealed that diabetic patients without previous MI have as high a risk of developing MI as nondiabetic patients with previous MI [141], although these findings have not been supported by all more recent, larger, studies [192,193]. It is possible that this effect is more pronounced in women than in men; thus studies comprising only men may not have reconfirmed this initial finding from a study where both sexes were included.

**Risk factors for CHD in diabetic patients**

Several studies have examined risk factors for CHD in T2DM patients, whereas little population-based prospective data are available on the risk and determinants of CHD in T1DM. The traditional risk factors, such as cigarette smoking, serum cholesterol, or BP, play an important role in the development of CHD in diabetic patients [172], but they explain only a part of the excess risk of ischemic heart lesions in diabetes [194–196]. The greater the number of coronary risk factors, the higher the susceptibility to CHD whether diabetic or not [140,166,172,173]. Hyperglycemia was found to be predictive of CHD death for T2DM in several studies [140,173,175,194,195]. The recently published results from the Nurses’ Health Study indicated duration of T2DM as an important factor associated with increased risk of death from all causes and fatal CHD [164]. Compared with those in nondiabetic persons, the multivariate relative risks of fatal CHD across categories of diabetes duration (5, 6–10, 11–15, 16–25, >25 years) were 2.75, 3.63, 5.51, 6.38, and 11.9 (p < 0.001 for trend), respectively. The combination of prior CHD and a long duration of clinical diabetes (i.e., >15 years) was associated with a 30-fold (95% CI 20.7–43.5) increased risk of fatal CHD [164]. These results need to be confirmed by other studies.

Serum cholesterol level has been found to be particularly important in predicting cardiovascular mortality [196]. In men with and without diabetes, CHD death rates increased markedly with increasing cholesterol levels. At every level of serum cholesterol, however, the CHD death rate was several times higher for diabetic than nondiabetic men, and the increase in absolute CHD mortality rate with higher serum cholesterol level tended to be disproportionately greater for diabetic men than for nondiabetic men [172].
The lipid profiles predictive of cardiovascular death were assessed in the Paris Prospective Study [197]. The relationship between baseline triglyceride levels alone was clear: individuals with the lowest baseline triglyceride levels had the lowest risk of death from CHD during 15 years of follow-up. Conversely, individuals with CHD had the highest triglyceride levels [197]. Also previous cross-sectional findings from the WHO Multinational Study suggested that triglycerides were a stronger predictor of CHD and PVD in several countries than was total cholesterol [198]. Adults with diabetes do not have total cholesterol or low-density lipoprotein (LDL) cholesterol concentrations that are significantly increased compared with the general population [199]. However, significantly more patients with diabetes have low HDL-C levels and elevated triglyceride levels compared with individuals without diabetes. Thus, lipid abnormalities, although not necessarily hypercholesterolemia, contribute to the prediction of CHD [140,173,174,197]. Elevated levels of triglyceride-rich lipoproteins lower HDL levels by promoting exchanges of cholesterol from HDL to very low density lipoprotein (VLDL) via cholesteryl ester transfer protein [200]. Diabetic persons with CHD more commonly have the combination of elevated triglycerides and low HDL than elevated total and LDL cholesterol levels [201].

Risk factors for CHD related to increased triglyceride concentrations include an increase in postprandial accumulation of remnant lipoproteins, which form highly atherogenic particles [202,203]. Erkelen's [204] has summarized available data on the altered composition of lipoproteins and lipids, so-called diabetic dyslipidemia, in patients with T2DM. Thus, diabetic dyslipidemia is characterized by (1) elevated levels of triglyceride; (2) normal levels of total cholesterol and LDL-C; (3) reduced levels of HDL-C; (4) elevated levels of apolipoprotein B; (5) a preponderance of small, dense LDL particles; and (6) increased levels of cholesterol-rich VLDL [204].

Central fat distribution appears to be one of the prominent indicators of CHD risk in the Paris, Gothenburg, and Framingham studies [197,205,206]. Also, android fat distribution independently of BMI contributes to the development of T2DM. This is particularly striking in women, as demonstrated in the Gothenburg Study [207]. Also in studies from Iowa [208] and Mauritius [209] the risk of developing diabetes increased with BMI as well as with the waist-to-hip ratio; the highest risk was observed for the highest combined tertiles. Björntorp [210] has suggested that the intra-abdominal tissue, particularly the portion drained by the portal circulation, has an exceedingly sensitive lipomobilization capacity. This results in high portal free fatty acid concentrations, which in turn generate through hepatic regulation hyperinsulinemia, hyperglycemia, and T2DM. This is associated with a cluster of metabolic abnormalities defined as the insulin resistance syndrome. To explain what makes fat accumulate intra-abdominally rather than subcutaneously, Björntorp suggested a number of possibilities that involve, besides genetic predisposition, endocrine perturbations, which could, for example, affect glucocorticoid and androgen levels. These perturbations can occur as a consequence of certain modern lifestyle factors such as stress, smoking, alcohol consumption, lack of physical activity, and the excessive intake of saturated fat.

Since the first suggestions of the insulin–atheroma relationship in the 1960s [211,212] there have been many clinical, experimental, and epidemiologic studies addressing this question. Insulin may directly promote atherosclerosis [213]. Although insulin correlates with a number of cardiovascular risk factors, such as BMI, body fat distribution, serum triglycerides, and HDL cholesterol, epidemiologic studies of the insulin–CHD relationship have provided inconsistent findings, both cross-sectionally and longitudinally. Moreover, populations that are clearly hyperinsulinemic such as Pima Indians, Naurans, Polynesians, etc., do not have a particularly high mortality from CHD [178,214–217]. Data from the Multiple Risk Factor Intervention Trial [218] demonstrated that insulin appears to contribute to cardiovascular risk in persons with apolipoprotein E 3/2 phenotype but does not explain the increased risk seen in these subjects. The authors suggested that this new finding, if confirmed, may throw further light on the role of insulin in atherosclerosis.

Long-term hyperglycemia may explain a large proportion of excessive cardiovascular mortality among T2DM patients not explained by classical coronary risk factors [140]. In addition, it was recently hypothesized that increased plasma and tissue glucose levels may be one important source of the increased oxidative stress seen in diabetic patients, which through various metabolic mechanisms could predispose to atheroma [73].

Albumin excretion rate has also been shown to be significantly associated with CHD morbidity after taking into account the confounding effects of raised BP and other coronary risk factors [219–221]. Also Nielsen et al. [222] has noted that the prevalence of CVD increases with increasing albumin excretion in T2DM. It was shown that the simultaneous occurrence of hyperinsulinemia and microalbuminuria is a strong predictor of CHD events in elderly nondiabetic people [223]. Several earlier studies performed among T1DM patients also observed a substantial increase in mortality from, and incidence of, CHD in those with nephropathy compared with T1DM patients without nephropathy [224,225]. In the Finnish cohort of childhood-onset T1DM patients the risk of CHD was high, about 3–4% after 20–33 years of follow-up when the patients were aged 19–49 years [158]. The risk of CHD was 10-fold in those who had developed nephropathy compared with those without nephropathy, and increased with increasing duration of nephropathy.

Prothrombotic and fibrinolytic factors act in concert to maintain hemostasis. Plasminogen activator inhibitor 1 (PAI-1), an enzyme needed to lyse a thrombus once formed, is synthesized in endothelial cells, vascular smooth muscle cells, adipose
Stroke and cerebrovascular diseases

Incidence and mortality

Diabetes mellitus magnifies the risk of cerebrovascular morbidity and mortality [162,234–241] (Table 69.1). The Framingham Study was one of the earliest to demonstrate an increased morbidity and mortality from cerebrovascular disease in patients with diabetes. Similar conclusions have been drawn from most, although not all, subsequent studies of patients with T2DM [235–240]. In a prospective study from Finland with a follow-up of 15 years, diabetes was the strongest single factor for stroke in a multivariate analysis (relative risk for men 3.4 and for women 4.9) [241]. Diabetes is a prominent risk factor for ischemic but not hemorrhagic stroke [241,256,257]. However, the recent joint report from the Honolulu Heart Program and the Framingham Study suggested that an increased risk of hemorrhagic stroke may also exist in T2DM [237].

Ischemic cerebrovascular disease accounts for about 80% of all strokes [258], but the female-to-male mortality ratio differs for stroke subtypes by ethnicity and age [259]. Diabetes mellitus is a definite independent risk factor for atherosclerosis and atherothrombotic brain infarction [240,247,260,261]. In the Framingham Study the incidence of atherothrombotic cerebral infarction for both sexes aged 45–74 years was greater by a factor of 2.5–3.5 in diabetic men that in non-diabetic subjects. Among the diabetic cohort, ischemic cerebral infarction comprised 88% of stroke events, 8% were subarachnoid or intracerebral hemorrhages and the rest of the strokes were of unknown type. Diabetes mellitus may also cause microatheroma in small vessels, such as lenticulostriate arteries, leading to lacunar stroke, one of the most common subtypes of ischemic stroke. Lacunar stroke is a unique subtype and requires specific clinical and radiological features for diagnosis. The presence of diabetes mellitus was associated only with symptomatic cerebral infarcts but not with silent infarcts [262], which are five times as prevalent as asymptomatic brain infarcts in the general population [263].

A negative correlation has been reported between diabetes, especially T1DM, and aneurysmal subarachnoid hemorrhage (SAH) [256]. While diabetes is one of the major risk factors for stroke in general, patients with aneurysmal SAH have a lower or equal prevalence of diabetes mellitus than does the general population. In particular, T1DM is rarely present prior to aneurysmal rupture [256]. However, there are also opposite opinions, since two studies claim that diabetes is closely associated with SAH and stroke [264,265]. Stroke patients with diabetes, or with hyperglycemia in the acute stage of stroke, have a higher mortality rate, worse neurologic outcome, and more severe disability than did those without [260–269]. The presence of diabetes, however, did not influence the outcome of parenchymatous hemorrhage or SAH [264,270]. There is insufficient information on the association of diabetes mellitus and radiologically confirmed SAH [271]. In a recent study, there was no difference in the presence of diabetes among SAH patients and controls [272]. Finland has one of the highest incidences of SAH [273]. During 1990–1998, 5630 first stroke events were registered in a community-based stroke register [274]. The subtype of the stroke was ischemic in 81%, hemorrhagic in 12%, and subarachnoid hemorrhage in 7% of the cases. The percentage of CT, MRI, angiography, or autopsy was 98%. The number of patients with diabetes was 1148 (20%). Of these, 19% were on insulin therapy and 81% were on diet alone or oral drug therapy. SAH was less common in patients with diabetes (1.0%) than patients without diabetes (8.2%, p < 0.00001).

There is much less information concerning the risk of stroke in T1DM. Deckert et al. [144] studying a group of patients who had had T1DM for more than 40 years, reported a 10% cumulative incidence of stroke and 7% mortality from stroke. The World Health Organization Multinational Study on...
of vascular disease in diabetes has indicated an overall raised cerebrovascular mortality in patients with T1DM, but with considerable variation between countries [275].

The data from the large nationwide cohort of Finnish childhood-onset T1DM patients showed that by the age of 50 years, and after 20–40 years duration of diabetes, the risk of stroke was equally high as that of acute CHD [158]. Also, the risk of stroke was equally common in both sexes. Previously, many T1DM patients who also had other risk factors for stroke (e.g. hypertension, smoking, dyslipidemia) had a high probability of dying relatively young from diabetic renal diseases or CHD, and so the number of T1DM patients surviving until “stroke risk age” (older than 65 years) has thus far been small. It is likely that improved treatment for hypertension and better blood glucose control will change this situation markedly, and stroke will become an important late complication of T1DM in the future.

### Risk factors for stroke in diabetic patients

In a Finnish study [276], fasting and 2-h glucose, a HbA1c duration of diabetes, and atrial fibrillation were the baseline variables predicting both fatal and nonfatal stroke events in T2DM patients during a 3.5-year follow-up. The association between duration of diabetes and risk of stroke was reported by these authors and also by other prospective studies from Finland [158,241]. The risk factors for stroke were the same for diabetic and nondiabetic subjects. Others have not found a significant relationship between ischemic stroke and diabetes duration, fasting glucose level, and serum HDL-cholesterol [277]. The Pima Indian Study did not find any relation between duration of T2DM and the risk of fatal stroke [278]. Results from the Honolulu Heart Program indicated that the serum glucose level at baseline was an independent risk factor for all cerebrovascular accidents and for thromboembolic strokes [237]. In a study from Oslo, nonfasting serum glucose was a risk factor for stroke mortality only in men with BMI above the median value, and only in univariate but not in multivariate analysis [254].

Several studies have provided sufficient evidence to indicate that hypertension and diabetes mellitus have direct effect on the occurrence of moderate to severe atherosclerotic stenosis of extracranial and intracranial arteries [279,280]. Furthermore some clinical series have suggested that patients with single lacunes may represent an entity different from those with multiple lacunes in that only the latter group have an increased frequency of arterial hypertension and diabetes mellitus [281,282]. However, in comparison with nonhemorrhagic lacunar stroke, patients with hemorrhagic lacunar stroke are less likely to have diabetes mellitus [283].

BP is a well-known risk factor for stroke in both types of diabetes [275], and the UKPDS trial suggested that controlling blood pressure in patients with T2DM reduced the stroke risk although not the risk of AMI [284]. Diabetic subjects have higher systolic BP than does the general population, particularly females [5,11,240,241,285]. It is suggested that diabetes may enhance the risk of stroke even in the absence of increased BP. In the nondiabetic population, there is a 40% increase in
stroke risk and a 25% increase in CHD risk with an elevation of 6 mmHg in excess of 75 mmHg diastolic BP. Diabetes increases the risk of both conditions by a factor of 2 or 3, and in diabetic subjects hypertension augments these risks [286]. Tuomilehto et al. [241] did not find evidence for an interaction between BMI, diabetes, and BP in a prospective study of stroke in a random population sample. There were neither additive nor multiplicative effects between diabetes and other CVD risk factors. Thus, it is unlikely that the risk of stroke associated with diabetes would be restricted to a particular subgroup, such as obese or hypertensive subjects. Moreover, it is possible that a part of the increased risk of stroke seen in hypertensive subjects is related to abnormalities of glucose metabolism, which are a common correlate of hypertension, particularly in the elderly [241,287,288].

It has been noted that the relative significance of risk factors for stroke varies with age and sex [289]. Hypertension, BP, smoking, hematocrit, and left-ventricular hypertrophy are strongly associated with stroke in patients younger than 65 years. Diabetes seems to be a stronger risk factor for stroke in females than in males [255,290]. Diabetes adversely affects cerebrovascular arterial circulation, akin to its effects in the coronary and lower extremity vasculature. Patients with diabetes have more severe extracranial atherosclerosis [291].

Besides being a risk factor for stroke, it has been shown that diabetes in the aftermath of the stroke event increases regional brain damage due to acute vascular insufficiency [292,293] and may have deleterious effects in both the short- and long-term prognosis of stroke [266–269,294]. It was shown that hyperglycemia directly increases regional brain damage due to acute vascular insufficiency [292]. Some studies have claimed, however, that neither a history of diabetes nor the glucose levels on admission are related to the outcome in an acute stroke event [295–300]. A number of studies have investigated whether diabetes has influence on the volume of damaged brain tissue in patients with stroke, as newer imaging methods have made the comparisons of stroke size possible. Most studies using computed tomography [267,268,300] and autopsy [301] have not found differences in the size of cerebral infarction between patients with and without diabetes mellitus.

There are at least three possible explanations for the worsening of stroke damage in the presence of hyperglycemia [302]. First, under the hypoxic conditions caused by a stroke, glucose is anaerobically metabolized to lactic acid, and the resultant cerebral intracellular and extracellular acidosis causes damage to neurons, glial, and vascular tissue. The effects of glycaemia may also differ at different anatomical sites. In the center of an infarct, excess glucose will worsen damage, whereas in the surrounding ischemic area, an increase in glucose level may lessen damage. Secondly, under hyperglycemic and hypoxic conditions the extracellular concentration of amino acids is increased because of excessive release combined with failure of energy-dependent re-uptake that would normally result in detoxification of glutamate and aspartate. A third theory is that with ischemia and hyperglycemia, together with neuronal hyperstimulation, intracellular calcium concentration increases, which may cause neuronal damage.

Excessive alcohol intake may increase the risk for stroke among patients with diabetes [303,304]. In addition, the Gothenburg population cohort study of mostly T2DM showed that alcohol combined with diabetes are important risk factors for cerebral infarction [252].

The increased incidence of and mortality from stroke in patients with T1DM is strongly related to the development of diabetic nephropathy [158,305]. The presence of nephropathy increases the stroke risk up to 10-fold compared with T1DM patients without nephropathy, and the stroke risk increases with increasing duration of nephropathy [158]. When proteinuria develops, not only is there an increase in systolic and diastolic BP but also an increase in von Willebrand factor, fibrinogen, platelet adhesiveness and number, and stimulated levels of tissue plasminogen activator and plasminogen activator inhibitor [305,306]. No evidence associates background diabetic retinopathy or proliferative retinopathy with an increased prevalence of stroke in T1DM [307]. Only one study has shown an increase in the relative risk of nonembolic ischemic stroke in diabetic subjects with retinopathy [308]. However, that study was unable to distinguish T1DM and T2DM subjects and duration of diabetes was not taken into account.

**Asymptomatic hyperglycemia and CVD risk**

It has been known for quite some time that individuals with overt diabetes have an increased risk for CVD, but the fact that subjects with asymptomatic diabetes or hyperglycemia are also at high risk for the development of CVD has been recognized only recently. Earlier studies such as the International Collaborative Group [309] in 1979 did not show consistent evidence for a threshold or graded association between asymptomatic hyperglycemia and CHD. In 1985, Epstein [310] reviewed 13 cohort studies that had studied postchallenge glucose and CHD; the results were conflicting.

Over the past decades, the prevalence of diabetes has increased markedly worldwide. Only around 1980, the National Diabetes Data Group (NDDG) [311] and WHO [312,313] formulated criteria for diagnosing diabetes based on glucose levels 2-h after a 75-g oral glucose load. The near-universal adoption of these diagnostic criteria has created order out of the confusion in the diagnostic criteria for diabetes and stimulated the interest in studying outcomes related to different categories of hyperglycemia. Now, after about two decades from the adoption of standard criteria for diabetes and glucose...
intolerance, epidemiologic evidence has been accumulated showing that asymptomatic diabetes, as well as hyperglycemia in the nondiabetic range was associated with the increased risk of CVD. Coutinho et al. recently made a meta-regression analysis of the data published between 1966 and 1996 from 20 studies comprising 95,783 individuals (94% male) followed on average for 12.4 years, and demonstrated that the progressive relationship between glucose levels and CVD risk extends below the diabetic threshold [314]. In data from the 23-year follow-up of the Paris Prospective Study of 7018 men, aged 44–55 years at baseline, Balkau et al. found that the relationship of fasting and 2-h post load glucose with mortality from all-causes and CHD were curvilinear, and that the increased risks already became apparent at the upper levels of the nondiabetic glucose distribution [315]. Results from a 22-year follow-up of nondiabetic Norwegian men (aged 40–59 years) showed a multivariate adjusted relative risk of 1.4 (95% CI 1.0–1.8) for CVD mortality in men with fasting whole blood glucose >4.7 mmol L\(^{-1}\) compared with <4.7 mmol L\(^{-1}\) [316]. An increased risk of developing CHD associated with the highest quartile of normal fasting glucose distribution was also observed in Australian nondiabetic women, but not in men [317].

Increasing evidence has been accumulated on the importance of hyperglycemia after an oral glucose challenge as an independent risk factor for CVD. In the Chicago Heart Study of 11,554 White men and 666 Black men aged 35–64 years, asymptomatic hyperglycemia (>11.1 mmol L\(^{-1}\)) 1 h after a 50-g glucose load showed increased CVD mortality in both White and Black men [318]. The 23-year follow-up of the Honolulu Heart Program among 8006 Japanese American middle-aged men also suggested an increasing risk gradient between 1-h postload glucose and CHD [319]. In the Hisayama Study, the multivariate adjusted relative risk for the incidence of CHD and stroke in Japanese aged 40–79 years was 1.9 (95% CI 1.2–3.2) for subjects with IGT compared with those who had a normal glucose tolerance [320].

In the late 1990s, a new surge of interest to study the relationship between asymptomatic hyperglycemia and the risk of CVD developed following the 1997 revision of diagnostic criteria for diabetes and milder degrees of hyperglycemia by the American Diabetes Association. In the Funagata diabetes study of a Japanese cohort aged 40 years or over, IGT but not impaired fasting glycemia (IFG) was associated with increased CVD mortality [321]. In the Hoorn study in the Netherlands in subjects with no prior history of diabetes, 2-h postload glucose proved to be a better risk predictor of CVD mortality than does fasting glucose or HbA1c [322]. Pooled longitudinal data from Mauritius, Fiji, and Nauru showed that isolated postchallenge hyperglycemia (>11.1 mmol L\(^{-1}\)) with a nondiabetic fasting plasma glucose (<7.0 mmol L\(^{-1}\)) doubled the CVD mortality risk [323]. Similar results were also observed in older Californian women [324]. A recent analysis of the data from US Second National Health and Nutrition Survey, including 3092 adults aged 30–74 years, revealed that subjects with IGT and undiagnosed diabetes defined by isolated postload hyperglycemia had 40% higher CVD mortality rates compared with those with a normal glucose tolerance [325]. Thus, there is no doubt that the risk of CVD starts to increase much before the development of fasting hyperglycemia. A 20-year follow-up of 117,629 women aged 30–55 years in the Nurses’ Health Study in the United States revealed that the risk of incident nonfatal MI and stroke started to increase 15 years before a clinical diagnosis of diabetes [326]. The multivariate adjusted relative risk for MI or stroke was 2.4, 3.2, and 3.6, respectively, for female nurses 15 years, 10–14.9 years, and <10 years before the onset of diabetes, in comparison with those who remained nondiabetic throughout the study.

The most convincing evidence of increased CVD risk related to asymptomatic hyperglycemia was provided by the DECODE (Diabetes Epidemiology: Collaborative Analysis of Diagnostic Criteria in Europe) Study. In this study, data from 10 prospective European cohort studies including 15,388 men and 7,126 women aged 30–89 years were collaboratively analyzed [327,328]. Multivariate Cox proportional hazard analyses showed that the risk for death from all-cause, CVD, CHD, and stroke significantly increased in subjects with asymptomatic diabetes defined according to either the fasting plasma glucose >7.0 mmol L\(^{-1}\) or the 2-h postload plasma glucose >11.1 mmol L\(^{-1}\). People with IGT also had an increased risk of death compared with those whose glucose tolerance was normal, but subjects with IFG did not have increased risk of death [327,328] (Table 69.2). Furthermore, the DECODE Study revealed that elevated 2-h glucose was a better predictor of mortality from all-cause, CVD, and CHD than did elevated fasting glucose alone [327,328]. A high 2-h glucose concentration was found to be associated with an increased risk of death, independent of the level of fasting blood glucose, whereas mortality associated with the fasting glucose concentration depended on the simultaneous level of 2-h glucose [327] (Table 69.3). Addition of 2-h glucose to a model based on fasting glucose improved the prediction (all p values <0.05, except for stroke). Conversely, fasting glucose did not add any further information once the 2-h glucose concentration was in the model (all p values >0.10) [327]. The largest absolute number of excess deaths was observed in subjects with IGT, especially those with IGT but normal fasting glucose, reflecting the fact that this category is the most prevalent abnormality in glucose regulation [327,328].

The observations from the DECODE Study on mortality related to 2-h glucose versus fasting glucose have also been extended to the prediction of the risk of serious incident CHD events (CHD death and nonfatal MI) based on five Finnish DECODE Study cohorts [329]. These results confirmed that in subjects without a prior history of diabetes the association of 2-h glucose with CHD incidence is graded and independent,
Table 69.2 Multivariate adjusted hazard ratios for deaths from cardiovascular diseases (CVD), coronary heart diseases (CHD), stroke, and all-cause according to the fasting and the 2-h glucose criteria: the DECODE Study

<table>
<thead>
<tr>
<th>Plasma (whole blood) glucose categories (mmol L⁻¹), subjects not known as diabetic</th>
<th>Fasting glucose criteria*</th>
<th>2-h glucose criteria†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IFG 6.1–6.9 (5.6–6.0)</td>
<td>Diabetes ≥7.0 (6.1)</td>
</tr>
<tr>
<td>CVD</td>
<td>1.09 (0.90–1.30)</td>
<td>1.48 (1.15–1.91)</td>
</tr>
<tr>
<td>CHD</td>
<td>1.07 (0.83–1.39)</td>
<td>1.43 (1.02–2.02)</td>
</tr>
<tr>
<td>Stroke</td>
<td>1.04 (0.70–1.56)</td>
<td>1.92 (1.16–3.17)</td>
</tr>
<tr>
<td>All-cause</td>
<td>1.11 (1.00–1.23)</td>
<td>1.65 (1.43–1.91)</td>
</tr>
<tr>
<td></td>
<td>7.8–11.0 (6.7–9.9)</td>
<td>Diabetes ≥11.1 (10.0)</td>
</tr>
<tr>
<td>CVD</td>
<td>1.34 (1.14–1.57)</td>
<td>1.55 (1.20–2.01)</td>
</tr>
<tr>
<td>CHD</td>
<td>1.28 (1.02–1.59)</td>
<td>1.64 (1.18–2.28)</td>
</tr>
<tr>
<td>Stroke</td>
<td>1.26 (0.88–1.79)</td>
<td>1.74 (1.01–2.99)</td>
</tr>
<tr>
<td>All-cause</td>
<td>1.40 (1.27–1.54)</td>
<td>1.92 (1.66–2.22)</td>
</tr>
<tr>
<td></td>
<td>11.1 (10.0)</td>
<td>Known diabetes‡</td>
</tr>
<tr>
<td>CVD</td>
<td>1.96 (1.62–2.37)</td>
<td></td>
</tr>
<tr>
<td>CHD</td>
<td>1.94 (1.51–2.50)</td>
<td></td>
</tr>
<tr>
<td>Stroke</td>
<td>1.72 (1.11–2.66)</td>
<td></td>
</tr>
<tr>
<td>All-cause</td>
<td>1.81 (1.60–2.05)</td>
<td></td>
</tr>
</tbody>
</table>

Note: Data are given as hazards ratios (95% confidence intervals), adjusted for age, center, total cholesterol, body mass index, systolic blood pressure, smoking, and sex.

*Using fasting plasma (whole blood) glucose <6.1 (5.6) mmol L⁻¹ as reference group.
†Using 2-h postload plasma (whole blood) glucose <7.8 (6.7) mmol L⁻¹ as reference group.
‡Compared with the models in Table 69.1, with only the 2-h glucose criteria, 2 df.
§Compared with the models in Table 69.1, with only the fasting glucose criteria, 2 df.

Table 69.3 Adjusted hazards ratios for death from cardiovascular diseases (CVD), coronary heart diseases (CHD), stroke, and all-cause with both the fasting and the 2-h glucose classes in the same model: the DECODE Study

<table>
<thead>
<tr>
<th>Plasma (whole blood) glucose categories (mmol L⁻¹), subjects not known as diabetic</th>
<th>Fasting glucose criteria*</th>
<th>2-h glucose criteria†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IFG 6.1–6.9 (5.6–6.0)</td>
<td>Diabetes ≥7.0 (6.1)</td>
</tr>
<tr>
<td>CVD</td>
<td>1.01 (0.84–1.22)</td>
<td>1.20 (0.88–1.64)</td>
</tr>
<tr>
<td>CHD</td>
<td>1.01 (0.77–1.31)</td>
<td>1.09 (0.71–1.67)</td>
</tr>
<tr>
<td>Stroke</td>
<td>1.00 (0.66–1.51)</td>
<td>1.64 (0.88–3.07)</td>
</tr>
<tr>
<td>All-cause</td>
<td>1.03 (0.93–1.14)</td>
<td>1.21 (1.01–1.44)</td>
</tr>
<tr>
<td></td>
<td>7.8–11.0 (6.7–9.9)</td>
<td>Diabetes ≥11.1 (10.0)</td>
</tr>
<tr>
<td>CVD</td>
<td>1.32 (1.12–1.56)</td>
<td>1.40 (1.02–1.92)</td>
</tr>
<tr>
<td>CHD</td>
<td>1.27 (1.01–1.58)</td>
<td>1.56 (1.03–2.36)</td>
</tr>
<tr>
<td>Stroke</td>
<td>1.21 (0.84–1.74)</td>
<td>1.29 (0.66–2.54)</td>
</tr>
<tr>
<td>All-cause</td>
<td>1.37 (1.25–1.51)</td>
<td>1.73 (1.45–2.06)</td>
</tr>
<tr>
<td></td>
<td>11.1 (10.0)</td>
<td>Known diabetes‡</td>
</tr>
<tr>
<td>CVD</td>
<td>1.96 (1.62–2.37)</td>
<td></td>
</tr>
<tr>
<td>CHD</td>
<td>1.94 (1.51–2.50)</td>
<td></td>
</tr>
<tr>
<td>Stroke</td>
<td>1.73 (1.12–2.68)</td>
<td></td>
</tr>
<tr>
<td>All-cause</td>
<td>1.82 (1.60–2.06)</td>
<td></td>
</tr>
</tbody>
</table>

Note: Data are given as hazards ratios (95% confidence intervals), adjusted for age, sex, center, total cholesterol, body mass index, systolic blood pressure, and smoking.

*Using fasting plasma (whole blood) glucose <6.1 (5.6) mmol L⁻¹ as reference group.
†Using 2-h postload plasma (whole blood) glucose <7.8 (6.7) mmol L⁻¹ as reference group.
‡Compared with the models in Table 69.1, with only the 2-h glucose criteria, 2 df.
§Compared with the models in Table 69.1, with only the fasting glucose criteria, 2 df.

Figure 69.1 Cumulative survival curves for incidence of coronary heart disease (a) and for cardiovascular mortality (b) according to the glucose categories defined by both fasting and 2-h plasma glucose criteria. The cumulative survival are estimated from Cox proportional hazards models and adjusted for age, cohorts, sex, body mass index, blood pressure, serum cholesterol, and smoking. Normal, fasting plasma glucose <6.1 mmol L⁻¹ and 2-h plasma glucose <7.8 mmol L⁻¹; IFG, impaired fasting glycemia only; IGT, impaired glucose tolerance only; IGR, impaired glucose regulation (IFG and IGT); DMF, undiagnosed diabetes by fasting glucose criteria alone (fasting plasma glucose ≥7.0 mmol L⁻¹ and 2-h plasma glucose <11.1 mmol L⁻¹); DMP, undiagnosed diabetes by 2-h postload glucose criteria alone (2-h plasma glucose ≥11.1 mmol L⁻¹ and fasting plasma glucose <7.0 mmol L⁻¹); DMC, undiagnosed diabetes by both fasting and 2-h glucose criteria combined (fasting plasma glucose ≥7.0 mmol L⁻¹ and 2-h plasma glucose ≥11.1 mmol L⁻¹); known DM, previously diagnosed diabetes. (For a color version of this figure, please see color plate section.)
and that 2-h glucose is superior to fasting glucose in assessing the risk of future CHD events (Figure 69.1). Thus, the DECODE Study unequivocally confirmed that asymptomatic postchallenge hyperglycemia independently increases cardiovascular morbidity and mortality and is a better predictor of the events than fasting hyperglycemia.

Whether postchallenge hyperglycemia is a true risk factor for CHD or merely a reflection of a high-risk phenotype remains to be determined [330]. Postchallenge plasma glucose concentrations are more strongly correlated with a variety of risk factors than are those of fasting plasma glucose in nondiabetic subjects, including central obesity, low-grade inflammation, and endothelial dysfunction [331]. It also remains unclear as to whether postchallenge hyperglycemia and postprandial hyperglycemia are similar or distinct characteristics. One study has suggested that mild-to-moderate mealtime hyperglycemia is an independent risk factor for increased IMT [332]. In a cohort at risk for diabetes, postchallenge glucose and glycemic excursions were found to be more strongly associated with carotid IMT than fasting glucose or HbA1c level [333]. More recently, Ceriello et al. has shown an independent and cumulative deleterious effect of postprandial hypertriglyceridemia and hyperglycemia on endothelial function and suggested that oxidative stress is the common mediator through which they exert such effects [87].

These findings may be important especially in the elderly population because postchallenge glucose levels increase to a much greater extent with advancing age than fasting glucose [324,334]. In the DECODE population, the prevalence of isolated postload hyperglycemia, that is, 2-h plasma glucose ≥11.1 mmol L^{-1} and fasting plasma glucose <7.0 mmol L^{-1}, was 0.7% in people 49 years or younger and 4.6% in those older than 70 years. In contrast, the prevalence of isolated fasting hyperglycemia, that is, 2-h plasma glucose <11.1 mmol L^{-1} and fasting plasma glucose ≥7.0 mmol L^{-1}, in these two age groups was 1.3 and 2.3%, respectively [335]. Fasting glucose levels in this aspect can be compared with diastolic blood pressure that beyond middle age does not rise with aging, whereas systolic blood pressure does and also is a stronger predictor of cardiovascular events [336]. With conclusive evidence, the guidelines for the treatment of hypertension have been modified to emphasize the importance of systolic blood pressure [14]. These recommendations are, however, more firmly based on intervention studies, and not only on epidemiologic evidence. The need for similar reassessment regarding postchallenge or postprandial versus fasting glucose, and for evaluating possible benefit of treatments focused on postprandial hyperglycemia, is required, especially in the elderly population. The guidelines on diabetes, prediabetes and cardiovascular disease published by the European Society of Cardiology [337] also emphasize the importance of testing all patients with CVD with an oral glucose tolerance test in order to detect people who have previously undiagnosed diabetes.

REFERENCES


Chapter 69

Rocchini AP: Insulin resistance and blood pressure regulation


Epidemiology of macrovascular disease and hypertension in diabetes mellitus

1023


229 Ridker PM, Stampfer MJ, Rifai N: Novel risk factors for systemic atherosclerosis: a comparison of C-reactive protein, fibrinogen, homocysteine, lipoprotein(a), and standard cholesterol screening as predictors of peripheral arterial disease. JAMA 2001;285:2481–2485.


Atherogenesis, coronary heart disease and insulin resistance syndrome in diabetes

Barak Zafrir and Jorge Plutzky
Brigham and Women’s Hospital, Boston, MA, USA

Key points

• Type 2 diabetes (T2DM), the insulin resistance that typically accompanies T2DM, and the constellation of abnormalities that define the metabolic syndrome all promote atherosclerosis and its complications.

• Significant molecular, cellular and clinical overlap exists in terms of the pro-atherosclerotic forces found in T2DM and the metabolic syndrome, including pathways such as inflammation.

• Both T2DM and atherosclerosis involve long subclinical phases prior to the manifestation of clinical disease, with attention focused on mediators and markers that may allow earlier identification and intervention. The metabolic syndrome, comprised of risk factors for developing diabetes and cardiovascular disease (CVD), can be considered as one such tool although its practical utility and meaning as a diagnostic and management tool remain controversial. Prediabetes as a condition may overlap with the metabolic syndrome.

• Specific mechanisms that account for increased cardiovascular risk in T2DM remain unclear, as does whether the nature of atherosclerosis in T2DM is unique as compared to nondiabetes.

• Hyperglycemia exerts pathologic effects on vascular tissues by multiple complex pathways including oxidative stress, increased inflammation, lipotoxicity involving free fatty acids, glycation of proteins, and cellular insulin resistance to name a few; specific mediators involved in these responses have been identified and continue to be uncovered. There is a convergence of most if not all these mechanisms on increasing inflammation, promoting fibrosis and thrombosis, and accelerating atherosclerosis, with effects on endothelial cells, vascular smooth muscle cells, monocyte/macrophages, lymphocytes and platelets.

• Recent large clinical trials have not demonstrated clear benefits of tight glycemic control in preventing CVD outcomes. Long-term follow-up studies suggest that optimal glycemic control may confer a long-term CVD benefit in diabetes.

• Atherogenic dyslipidemia together with ectopic lipid accumulation in tissues, negatively affect glucose metabolism and insulin signaling in diabetes.

• Obesity, and in particular visceral adiposity, may increase inflammation and promote diabetes and CVD through endocrine and paracrine effects.

Diabetes mellitus and the burden of cardiovascular disease

The two- to fourfold increased risk of cardiovascular disease (CVD) in diabetes is unequivocal. Indeed, one could argue that the way in which we define diabetes—hyperglycemia—is only a marker for the major health threat in diabetes, which is CVD [1,2]. Cardiovascular (CV) events remain the major cause of death for those with diabetes, accounting for about 70% of all deaths [3]. Furthermore, the 7-year incidence of myocardial infarction (MI) in patients with type 2 diabetes (T2DM) and no history of MI may equal the incidence of patients with a prior MI but no diabetes [4]. Particularly concerning is the increasing prevalence of T2DM in younger patients, raising concerns about the future CV risk in such patients. Despite this, as discussed elsewhere, recent clinical trials fail to demonstrate the ability to decrease CV events by improving glucose control, a possible consequence of multiple issues. These complexities compel ongoing examination of the pathologic forces that promote diabetes and atherosclerosis in patients with diabetes, the focus of this chapter. Although the focus of attention in this area is inevitably coronary atherosclerosis and MI, it is important to note that diabetes increases the risk of atherosclerosis systemically, with increased peripheral arterial disease and cerebrovascular disease, worsens outcomes with CV interventions, and worsens outcomes with congestive heart failure. In considering the state of knowledge around the critical issues of the relationship between T2DM and CVD, key central challenges become evident. The first is the overlap between T2DM and CVD: both diseases are quite common—making causal separation difficult; can manifest themselves in distinct ways—sugesting oversimplification with our current diagnostic groups; have long preclinical phases—raising the prospect of earlier intervention; and appear to share some pathogenic mechanisms—again posing challenges in elucidating drivers of disease and therapeutic strategies. Such challenges can only be
met by further considering our current understanding of the disease processes themselves.

**Insulin resistance (metabolic) syndrome**

Patients with T2DM typically have an inadequate response to a given concentration of insulin as seen under controlled study conditions, so-called insulin resistance. It has been estimated that a third of the US population meet the criteria for the metabolic syndrome [5]. Early observations noted that insulin resistance, even before the presence of frank hyperglycemia, was often associated with other clinical conditions including central obesity, glucose intolerance, increased triglycerides, decreased levels of high-density lipoproteins, and hypertension. Importantly, all of these components of the metabolic syndrome also promote atherosclerosis. The constellation of these abnormalities was referred to as the metabolic syndrome (also called insulin resistance syndrome, syndrome X, Reaven’s syndrome, among other names). Several lines of evidence specifically associated this clustering of cardiometabolic risk with visceral adiposity, as reflected by waist circumference. The clinical relevance of the metabolic syndrome was derived from data suggesting that the presence of the syndrome identified a significantly increased risk for future T2DM and CVD. Countering opinions suggested that the presence of metabolic syndrome offered no clinical utility, and that the future risk of CVD was no different than the sum of the individual metabolic syndrome components present in a given patient.

Despite these valid points, in addition to offering a convenient, appealing shorthand to clinicians who commonly encountered this risk cluster, discussions around the metabolic syndrome also provided a construct for studies into the nature and progression of atherosclerosis and cardiovascular complications in T2DM and those at increased risk for future diabetes. Subsequent attention has focused on clinical groups that overlap to various extents with the metabolic syndrome, including prediabetes (A1C ≥ 5.7%, <6.5%), impaired glucose tolerance, and impaired fasting glucose. These clinical designations may identify patients earlier in the natural history of T2DM and atherosclerosis, allowing potential inputs to diabetes and atherosclerosis to be isolated.

Insulin resistance and its associated abnormalities are all risk factors for the development of both diabetes and CVD. In response to insulin resistance, pancreatic β cells secrete increasing amounts of insulin, resulting in hyperinsulinemia. Whether or not hyperinsulinemia may contribute to atherosclerosis in patients with diabetes remains debated even as more evidence has argued against pharmacologic insulin as promoting atherosclerosis. The presence of hypertension in insulin resistance and T2DM is noteworthy, suggesting systemic changes in the vasculature and that mechanical forces found in hypertension, like shear stress, may already be priming the endothelium for pathogenesis. Several clinical observations are particularly intriguing in this regard. First, some relatively large studies find that when more sensitive screens for T2DM, like oral glucose tolerance testing, are done in nondiabetic patients, a significantly higher incidence of diabetes is uncovered among those with hypertension than among otherwise matched non-hypertensive subjects. Second, various studies using different means suggest that altered endothelial vasoreactivity may be among the first abnormalities in subjects destined for future T2DM. Preclinical studies support the notion that endothelial dysfunction, discussed elsewhere, may involve not just vasomotor function and hypertension but also glucose and fatty acid handling, key abnormalities in diabetes and atherosclerosis.

Insulin resistance syndromes are characterized by atherogenic dyslipidemia involving increased triglyceride levels (TG), low concentrations of high-density lipoprotein cholesterol (HDL-C), and smaller, denser low-density lipoprotein cholesterol (LDL-C). For a given LDL level, patients with smaller, denser LDL have increased levels of apolipoprotein B (apoB), which is more prone to oxidation and retention within the vessel wall. Oxidized LDL is considered a key element in foam cell formation and atherosclerosis. Hypertriglyceridemia in diabetes derives from increased secretion of TG-rich VLDL as well as impaired VLDL hydrolysis by enzymes like lipoprotein lipase (LPL). LPL has been reported to generate endogenous PPARα ligands, a potentially protective effect that may be lost in T2DM.

In attempting to identify the core defect in the insulin resistance syndrome, several approaches have implicated central obesity as a key culprit. Adipose tissue is now recognized as a dynamic, well-vascularized organ exerting endocrine, paracrine, and autocrine effects. Different depots of fat have distinct functional effects. Visceral adiposity is associated with increased inflammation and other abnormalities associated with diabetes and atherosclerosis [6]. Interestingly, some of the benefits seen with pioglitazone as a thiazolidinedione may involve expansion of subcutaneous fat, as a less dangerous depot, and contraction of visceral fat. Brown fat has been documented to be present in adult humans and may limit adiposity through thermogenesis. Especially relevant for this discussion are findings that report perivascular fat as brown-like in nature. Mechanisms linking adiposity to atherosclerosis in diabetes include increased free fatty acid (FFA) levels, increased cytokine release, and alterations in adipokines, for example adiponectin, as an anti-inflammatory protein produced more by subcutaneous than visceral fat, whose plasma levels correlate inversely with CV risk [7]. Recent studies implicate resident inflammatory cells in adipose tissue as promoting inflammation and diabetes. Clinical studies suggest that low-grade inflammation may predict future T2DM as well as CVD, further underscoring shared pathways of pathogenesis.

Insulin resistance syndromes are also thought to involve a more prothrombotic state. Shifts in the fibrinolytic balance have been reported, including increased plasminogen activator inhibitor 1 (PAI-1), which is made by both adipocytes and
endothelial cells, and decreased levels of the endogenous anticoagulant total plasminogen activator (tPA). Platelets also appear hyperreactive in insulin resistance and T2DM. Insulin resistance and T2DM can often have significant hepatic changes, influencing dyslipidemia and coagulation; increased deposition of hepatic fat may be part of this process.

Despite disappointing studies that failed to show decreased CV events through better control of glucose, one cannot ignore the extensive database identifying hyperglycemia as a pathogenic mechanism in either T2DM or insulin resistance. Increased glucose levels induce multiple cellular responses linked to inflammation and atherogenesis, as discussed further later. In fact, linear relationships between hyperglycemia and CV risk are evident even at levels of glucose and A1c that are considered within a normal range. Such relationships may be particularly relevant in prediabetes, impaired glucose tolerance/fasting glucose, and metabolic syndrome in which patients, and their vasculature, are exposed to hyperglycemia after meals, but can still achieve glucose levels after fasting or A1c values over time that are still in a "nondiabetic" range.

Putting this picture together, T2DM and the insulin resistance that typically accompanies it leads to a complex interplay of multiple clinical conditions that are generally unified in promoting inflammation and atherosclerosis. This pattern extends to earlier phases of diabetes, including prediabetes, impaired fasting glucose, and impaired glucose tolerance, as well as insulin resistance/metabolic syndrome. A central question regarding this focuses on early states associated with future risk of diabetes and cardiovascular disease. Definitions and criteria of the metabolic syndrome were revised several times by different medical associations, which reflect the disagreements and varied opinions regarding the etiology and nature of this syndrome. The criteria that make up the metabolic syndrome are listed in Table 70.1, with the syndrome considered present if any three of the five variables are found. Other groups have proposed other criteria that differ in various subtle if not semantic ways (Table 70.1). Certain aspects of the metabolic syndrome have also been defined differently for specific groups, like waist circumference as a surrogate for visceral adiposity among Southeast Asians, who may have more visceral fat at lower waist circumference.

Considerable controversy arose over the value of the metabolic syndrome and the rigor with which it had been established. Multiple prospective studies suggested that the presence of the metabolic syndrome was associated with an increased risk for future CVD [8]. However, conflicting opinions exist as to whether the presence of the syndrome confers a greater CVD risk than the risk associated with the individual components of the metabolic syndrome. Other criticisms include the medical value of diagnosing the metabolic syndrome if it will not lead to any alteration in management [9,10]. A further debate is whether the metabolic syndrome is a collection of unrelated risk factors, or combined by a common pathogenic mechanism. In 2010, the WHO Expert Consultation cast the metabolic syndrome as an educational concept that focuses attention on complex multifactorial health problems, and recognized it as a premorbid condition rather than a clinical diagnosis, with limited practical utility as a diagnostic or management tool [11].

Although the pathogenesis of metabolic syndrome remains debated, insulin resistance—as formally defined in insulin clamp studies—has been proposed as a key link between the components of the metabolic syndrome, thus explaining the use of the term “insulin resistance syndrome” by some [12]. Given the constellation of abnormalities that seem to track together with insulin resistance, efforts have been made to identify the core defects, whether mechanistic or clinical, that account for the syndrome. Clinically, significant evidence would point to adiposity, and in particular visceral adiposity, as key driver on insulin resistance. Genetic predispositions may interact with environmental changes such as excessive caloric intake, decreased physical inactivity, and overall impaired energy balance that fosters increased storage of fatty acids to be stored in adipocytes as TG. This process may promote TG accumulation in ectopic locations, like visceral fat, liver and skeletal muscle [13,14]. This dysregulated energy balance is also associated with abnormalities in lipolysis, increased FFA, and increased specific molecular species like ceramides that have also been associated with insulin resistance and its complications [15]. The antilipolytic effects of insulin, which is a sensitive parameter of insulin action, may be impaired in insulin resistance states, leading to an increased rate of lipolysis and increased release of FFA in the circulation.

Another related core feature in insulin resistance appears to be inflammation, and in particular inflammation in visceral fat. Considerable interest has focused on accumulation of monocytes/macrophages in visceral fat where they may play a paracrine and endocrine role driving systemic inflammation and cardiometabolic complications. Consistent with this line of thinking are investigations testing the use of salicylates, as anti-inflammatory reagents, to treat diabetes [16].

**Hyperglycemia, glycemic control, and CVD**

Diabetes is associated with atherosclerotic CVD, with a greater extent of disease and more complicated consequences than in the nondiabetic population. Hyperglycemia defines the diagnosis of diabetes, and confers increased risk of vascular morbidity and mortality in a wide range of epidemiologic and cohort studies [17]. Meta-analyses have further shown that chronic hyperglycemia and elevated glycosylated hemoglobin are associated with an increased risk for CV events and all-cause mortality among patients with T2DM, likely independently from other conventional risk factors [18,19]. Interestingly, the risk relationship between hyperglycemia and measures like A1c may be present at levels well below the threshold considered diagnostic of T2DM. Nevertheless, despite this, three recent
Blood pressure

Other

Microalbuminuria: urinary excretion rate of >20 mcg min⁻¹ or albumin : creatinine ratio of >20 mg g⁻¹

Table 70.1 Criteria for clinical diagnosis of metabolic syndrome

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Definition</td>
<td>IGT, IFG, T2DM, or insulin resistance + any 2 of additional criteria below:</td>
<td>Any 3 of 5 of the criteria below:</td>
<td>IGT or IFG plus any of the following based on clinical judgment</td>
<td>Central obesity + any 2 of the other criteria below:</td>
</tr>
<tr>
<td>Glucose</td>
<td>IGT, IFG, or T2DM</td>
<td>FPG &gt; 110 mg dL⁻¹ (including diabetes)</td>
<td>IGT or IFG (but not diabetes)</td>
<td>FPG ≥ 100 mg dL⁻¹ (including diabetes)</td>
</tr>
<tr>
<td>Body weight</td>
<td>WHR &gt; 0.9 in men, &gt;0.85 in women, and/or BMI &gt; 30 kg m⁻²</td>
<td>Men: WC ≥ 102 cm; Women: WC ≥ 88 cm</td>
<td>BMI ≥ 25 kg m⁻²</td>
<td>Increased WC (ethnicity specific)</td>
</tr>
<tr>
<td>Lipids</td>
<td>TGs ≥ 150 mg dL⁻¹ and/or HDL-C &lt; 35 mg dL⁻¹ in men or &lt; 39 mg dL⁻¹ in women</td>
<td>TGs ≥ 150 mg dL⁻¹; HDL-C &lt; 40 mg dL⁻¹ in men or &lt; 50 mg dL⁻¹ in women</td>
<td>TGs ≥ 150 mg dL⁻¹; HDL-C &lt; 40 mg dL⁻¹ in men or &lt; 50 mg dL⁻¹ in women; on drug Rx</td>
<td>TGs ≥ 150 mg dL⁻¹; HDL-C &lt; 40 mg dL⁻¹ in men or &lt; 50 mg dL⁻¹ in women; on drug Rx</td>
</tr>
<tr>
<td>Blood pressure</td>
<td>≥ 160/90 mmHg</td>
<td>≥ 130/85 mmHg</td>
<td>≥ 130/85 mmHg</td>
<td>≥ 130/85 mmHg or on drug Rx</td>
</tr>
<tr>
<td>Other</td>
<td>Other features of insulin resistance</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

a Insulin sensitivity measured under hyperinsulinemic euglycemic conditions, glucose uptake below lowest quartile for background population under investigation.

b Europids, 94 cm in men and 80 cm in women (in the USA, the ATP III values are 102 cm male, 88 cm female); South Asian and Chinese, 90 cm in men and 80 cm in women; Japanese, 85 cm in men and 90 cm in women; South and Central America, South Asian recommendations until more specific data become available; Sub-Saharan Africa, Eastern Mediterranean and Middle East populations, European data until more specific data becomes available.

c It is recommended that the IDF cut points be used for non-Europeans and either the IDF or AHA/NHLBI cut points used for people of European origin until more data are available.

AACE, American Association of Clinical Endocrinologists; AHA, American Heart Association; BMI, body mass index; BP, blood pressure; FPG, fasting plasma glucose; HDL-C, high-density lipoprotein cholesterol; IAS, International Atherosclerosis Society; IASO, International Association for the Study of Obesity; IDF, International Diabetes Federation; IFG, impaired fasting glucose; IGT, impaired glucose tolerance; NCEP-ATP III, National Cholesterol Education Program Adult Treatment Panel III; NHLBI, National Heart, Lung, and Blood Institute; T2DM, type 2 diabetes mellitus; TG, triglyceride; WC, waist circumference; WHF, World Heart Federation; WHO, World Health Organization; WHR, waist-to-hip ratio.

major clinical trials — Action to Control Cardiovascular Risk in Diabetes (ACCORD), Action in Diabetes and Vascular Disease (ADVANCE), and the Veterans Affairs Diabetes Trial (VADT) — all failed to demonstrate clear benefits of aggressive glycemic control in decreasing CVD events [20–22]. That is in contrast to the clear benefit displayed in past trials demonstrating conclusively that microvascular complications can be reduced by better glycemic control in both patients with type 1 and type 2 diabetes [23–26]. Intense interest has focused on this “glucose paradox” in which hyperglycemia increases CV risk but lowering glucose fails to decrease CV events, at least to the extent that might have been predicted. Some antidiabetic agents may increase CV risk, offsetting any potential benefit in CV risk reduction. More aggressive glucose control, especially with agents like insulin, often brings hypoglycemia, which may increase CV events. This may be particularly an issue in older patients. The increased use of other interventions that decrease CV risk, as would be utilized to treat the control group in a clinical trial, makes it increasingly harder to show benefit. Subtle variables like the duration of diabetes, and when treatments like statin use may have been implemented, may also be factors. Indeed, subgroup analyses do suggest that patients treated more aggressively for T2DM earlier in their natural history may experience better CV outcomes [27–29]. The results of several meta-analyses incorporating these recent studies demonstrated a benefit of improved glycemic control on nonfatal MI but no difference in overall mortality [30–32]. It should also be noted that longer follow-up may be necessary in order to see the benefit of diabetes therapies, as has been reported for significant CV risk reduction in T1DM in the DCCT/EDIC study and in T2DM patients in a UKPDS 10-year follow-up study [33,34]. Taken together, the case can be made that optimal glycemic control in patients with diabetes does confer a long-term CVD benefit even if management programs need to be tailored to individual patients and more information is needed along with better therapies [17,28].
Diabetes and atherosclerosis

Coronary artery disease in patients with diabetes is characterized by accelerated atherosclerosis occurring at an earlier age than the nondiabetic population. It is more extensive, diffuse in multiple arterial territories, and involves more distal and smaller coronary arteries. Angiography studies display fewer collateral vessels in diabetic patients compared with nondiabetics, which may result from failure of compensatory vasculogenesis in response to ischemia in diabetes [35]. Not infrequently these features limit and complicate revascularization possibilities. Plaque rupture and thrombus formation are more common in diabetes, leading to reduced outcomes after acute coronary syndromes, stroke, and heart failure.

Despite extensive examination of CVD in the context of diabetes, whether diabetic atherosclerosis differs fundamentally in terms of mechanisms from nondiabetic atherosclerosis or just an acceleration of the same processes remains poorly resolved. Atherosclerosis in diabetes is likely a multifactorial process [36]. Traditional coronary heart disease risk factors such as hypertension and dyslipidemia are prevalent in diabetes, influencing the risk of atherosclerosis by multiple mechanisms (Figure 70.1). Intravascular deposition of lipids and cellular debris characterize the atherosclerotic plaque formation. However, hyperglycemia, insulin resistance, and chronic inflammation are key processes promoting atherosclerosis development and progression in diabetes, by multiple mechanistic signaling pathways linked by common regulatory systems [37,38]. It is difficult to separate the effects of hyperglycemia from pro-atherogenic effects of other risk factors, especially dyslipidemia, in contributing to the accelerated atherosclerosis observed in diabetes.

In morphologic autopsy studies, atherosclerosis in diabetes is characterized by more extensive disease with distal involvement of coronary arteries as compared to nondiabetics, and increased plaque burden with higher rates of healed ruptured plaques observed. Larger necrotic cores along with inflammatory infiltrates of macrophages and T-cell lymphocytes have been reported in diabetic atherosclerotic vessels [39]. Intravascular ultrasound (IVUS) studies of atherosclerosis suggest that diabetes carries a greater baseline burden of coronary atherosclerosis with greater atheroma volume, smaller lumen dimensions, inadequate expansion of the vessel wall, and a more rapid progression [40]. Computed tomography (CT) imaging studies display higher coronary artery calcium scores in diabetes subjects without known coronary artery disease, than in nondiabetic individuals, with greater mortality rates. In addition, a higher rate of obstructive multivessel disease is seen in CT angiography of diabetic patients [41–43]. Overall, morphologic and imaging studies of diabetic atherosclerosis consistently demonstrate greater plaque burden, inflammation, and calcification than in nondiabetic individuals.

Similar to T2DM, T1DM is associated with early changes in vascular function, and an increased risk for premature coronary heart disease. Standard modifiable risk factors, such as smoking, lipids, and hypertension, enhance risk in T1DM, and renal disease with proteinuria, which can be found in T1DM, is independently predictive of coronary artery disease [44]. Multislice CT studies suggest that excess coronary artery calcification is seen more often in T1DM as compared to a relatively higher proportion of noncalcified plaques observed in patients with T2DM [45]. The fact that T1DM is less often associated with visceral adiposity suggests fundamental mechanistic inputs in the atherosclerosis that complicates T1DM and T2DM, even if final common pathways remain shared between these clinical entities. Atherosclerosis is initiated by the adhesion of monocytes to endothelial cells in the vascular wall, transmigrating into the subendothelium, and differentiating into macrophages [46]. Intravascular accumulation of lipids leads to LDL oxidation and induction of an inflammatory response. Macrophages, via uptake of lipids, transform into foam cells and accumulate in the vessel wall, resulting in fatty streak formation which constitutes the lipid core of the atherosclerotic plaque. During atherogenesis and atherosclerosis, inflammatory substances and growth factors are secreted by multiple sources, including endothelial cells, vascular smooth muscle cells (VSMCs), and inflammatory cells, stimulating infiltration of the intima by FFA, free fatty acids.
leukocytes and the migration and proliferation of VSMCs, which appear responsible for the formation of a collagen-rich fibrous cap [47]. Cell death and newly recognized processes such as autophagy contribute to the highly thrombotic necrotic core found in lesions. Ongoing accumulation of macrophages, lymphocytes, lipids, and connective tissue matrix, surrounded by a fibrous cap, represents the mature atherosclerotic plaque. Activated metalloproteinase enzymes further degenerate the collagen matrix, and the proliferation of vasa vasorum into the adventitia may lead to intraplaque hemorrhage, further promoting plaque necrosis and thinning of the fibrous cap. Given that the fibrous cap is what separates the circulation, with its highly active, efficient coagulation system, from the thrombogenic necrotic core, the stability of the fibrous cap becomes a critical factor that determines whether or not plaque rupture and myocardial infarction will occur. Features of instability and vulnerability of the plaque include a large lipid/necrotic core, a thin fibrous cap, plaque hemorrhage, and an inflammatory infiltrate rich in monocytes and macrophages in the shoulder region. New insight into monocyte biology and monocyte transition to macrophages has also revealed the existence of different monocyte subtypes, including ones that limit inflammation and atherosclerosis, and others that promote these processes. Hyperglycemia, FFA, and diabetes may shift macrophage subtypes towards those that are more pathogenic.

Insulin resistance and diabetes may further stimulate and induce many of the higher pathogenic aspects of atherosclerosis. Insulin resistance in macrophages enhances apoptosis and promotes formation of necrotic core in advanced atherosclerotic plaques [48]. Increased wall stress, as seen with hypertension, and the characteristics of blood-flow, as seen with increased shear stress, contributes to the heightened thrombogenicity of the vessel wall. In diabetes, a reduction in the strength of the fibrous cap has been reported, which may raise the likelihood of plaque rupture. Diabetes is also associated with a hypercoagulable state, with increased PAI-1 levels and decreased tPA, changes associated with increased propensity of clot formation. Platelets in diabetes also appear hyperfunctional [49].

**Pathogenesis linking metabolic syndrome, diabetes, and atherogenesis**

Multiple mechanisms likely foster atherosclerosis in diabetes. Hyperglycemia and insulin resistance mediate vasoconstrictive, proinflammatory, and prothrombotic vascular responses which promote pro-atherosclerotic events [17]. Atherosclerosis in diabetes is accelerated by endothelialial and VSMC dysfunction, immune and inflammatory activation, which are driven by metabolic derangements including hyperglycemia, impaired insulin signaling, and dyslipidemia with elevated FFA production. Various inputs to these responses will be considered here.

**Hyperglycemia and glucotoxicity**

The mechanisms through which hyperglycemia promotes the development of vascular disease remain incompletely understood. Hyperglycemia may promote atherosclerosis and CVD by altering hepatic and peripheral lipoprotein metabolism, resulting in changes such as increased triglycerides, lower HDL, and smaller, denser LDL that may be more prone to uptake in the vessel wall and oxidation. Hyperglycemia also facilitates the development of proteinuria and chronic kidney disease in patients with T2DM. The vascular wall, and in particular the microvasculature, is sensitive to changes in glucose concentration. Hyperglycemia exerts adverse effects on endothelial function, inhibiting the availability and activity of nitric oxide (NO), and impairing endothelium-dependent vasodilation in diabetes subjects [50]. Evidence from multiple experiments supports the view that hyperglycemia induces a range of pro-atherogenic effects, through several suggested mechanisms (Figure 70.2). The strength of such data only adds to the confusion over why treating hyperglycemia has not had more definitive effects on decreasing CV events.

**Oxidative stress and ROS**

Hyperglycemia enhances oxidative stress by inducing reactive oxidative species (ROS), promoting vascular dysfunction. Oxidative stress has a central role in the development of atherosclerosis through activation of several pathways involved in the pathogenesis of diabetic complications [51]. Through these pathways, oxidative stress promotes inactivation and reduced bioavailability of NO, increase in VSMC proliferation, macrophage adhesion, vascular apoptosis and remodeling, and enhances platelet activation. Oxidative stress can also lead to increased production of oxidized LDL in the vessel wall. Hyperglycemia induces overproduction of superoxide and NADPH oxidase, both a source of vascular oxidative stress in diabetes, which can damage vascular cells by activating and interacting with different pathogenic mechanisms as outlined.

**DAG production and PKC activation**

Hyperglycemia and changes in fatty acid oxidation increase the production of diacylglycerol (DAG) from glucose, thus increasing activation of the intracellular signaling molecule protein-kinase C (PKC). Activation of PKC isoforms phosphorylates various target proteins and affects vascular permeability, decreases NO production, and promotes extracellular matrix synthesis, cell growth, leukocyte adhesion, and cytokine activation [52]. PKC increases oxidative stress through activation of NADPH oxidase, decreasing NO production [53]. PKCs further impair insulin actions in the vasculature.

**Polyl and hexosamine pathways**

During hyperglycemia, endothelial cells are unable to reduce glucose uptake, resulting in glucose metabolism that is shifted to other pathways. Increased activity of aldose reductase and the polyl pathway due to high influx of glucose, results in
toxic accumulation of compounds such as sorbitol which generates ROS and promotes vasoconstriction [54]. Excess fatty acid oxidation may also stimulate diabetic complications by increasing the flux of fructose 6-phosphate into the hexosamine pathway [49–51]. The hexosamine pathway causes reversible modifications at regulatory phosphorylation sites on proteins involved in insulin signaling. Hexosamines may also mediate hyperglycemia-induced increase in gene transcription and activation of transforming growth factors (TGF) and PAI-1, as well as inhibit endothelial NO synthase (eNOS) activity in endothelial cells.

**Increased intracellular AGE formation**

Glucose-induced formation of nonenzymatic advanced glycation end products (AGE) have major importance in the pathogenesis of vascular dysfunction in diabetes. AGEs are formed through prolonged exposure of proteins and lipids to high concentrations of circulating glucose. Many different heterogeneous proteins, lipids and nucleic acids undergo irreversible post-translational modifications through glycation, with cross-linking of the proteins with reducing sugars. Indeed, the use of the A1c as a long-term measure of glucose control is simply a measure of the glycation of hemoglobin. Amino groups on proteins are bounded to glucose moieties due to hyperglycemia, and undergo nonenzymatic glycosylation, reduction, and Amadori rearrangements, eventually leading to the formation of AGE. AGE is implicated in the atherosclerotic process, either through their direct effects or by AGE interaction with the receptor for AGE (RAGE). High glucose levels and oxidative stress increase AGE–RAGE interactions, altering cellular signaling, promoting gene expression and enhancing leukocyte recruitment and the release of pro-inflammatory and adhesion molecules [55]. AGE generates ROS by stimulating NADPH oxidase activity, reducing eNOS activity and endothelial production of prostacyclin [56]. Levels of AGE in the serum of T2DM patients are negatively associated with endothelial vasodilatation, and AGE is associated with increased vascular stiffness. AGE and its receptor-ligands interactions activate NF-κB, a key proximal mediator of inflammatory responses, further increasing vasoconstrictive and proinflammatory responses. Accordingly, diabetic mouse models show that RAGE activation by pro-inflammatory protein ligands has a central role in the formation and progression of atherosclerotic lesions [57]. AGE–RAGE interactions also induce pro-coagulant forces in endothelial cells, increase expression of tissue factor by monocytes, activate inflammatory response,
stimulate pro-thrombotic pathways and an increase in matrix degrading enzymes associated with plaque rupture.

**NF-κB activation**

High glucose concentrations enhance the activation of NF-κB, which is a key pro-inflammatory nuclear transcription factor that controls multiple inflammatory targets. NF-κB activation by hyperglycemia increases expression of endothelial adhesion molecules, facilitating monocyte adhesion and vessel wall entry. Investigators have shown that transient hyperglycemia can induce epigenetic changes in a promoter region of NF-κB subunit in aortic endothelial cells, suggesting that transient hyperglycemia may cause persistent atherogenic effects during subsequent normoglycemia [58]. Hyperglycemia-induced activation of PKC has also been implicated in the activation of NF-κB in endothelial cells and VSMCs [59,60]. Studies have established links between activation of NF-κB, development of inflammatory phenotype, insulin resistance, and endothelial dysfunction. NF-κB inhibition may be a therapeutic target for treating diabetes and CV disease.

Overall, from a mechanistic perspective, hyperglycemia may induce chronic vascular complications via formation of toxic metabolites such as AGE, ROS, sorbitol, and persistent activation of PKC pathways, promoting an inflammatory milieu and accelerating atherosclerosis (Figure 70.2) [59–61]. Nonetheless, it is difficult to differentiate the relative contribution to the progression of atherosclerosis in diabetes from hyperglycemia itself versus other effects of dyslipidemia and other CVD risk factors often found in association with hyperglycemia and diabetes.

**Insulin resistance and hyperinsulinemia**

Insulin resistance has been associated with increased risk of coronary artery disease, and may promote atherosclerosis as an independent risk factor [8,62,63]. Insulin resistance can derive from defective insulin receptor signaling or overstimulation of insulin receptor pathways caused by hyperinsulinemia. The formal definition of insulin resistance involves demonstration of the need for higher infusion rates of insulin in order to maintain glucose at a certain level, thus establishing a cellular level of “resistance” to insulin. After binding to its cognate cell surface receptors, insulin activates the insulin receptor tyrosine kinase that phosphorylates insulin receptor substrates (IRS), resulting in stimulation of two divergent pathways (Figure 70.3): phosphatidylinositol-3 kinase (PI-3K)-dependent pathway, which mediates metabolic and anti-apoptotic effects, and the mitogen-activated protein kinase (MAP-K) which is linked to proliferative and mitogenic effects promoting cell growth and proliferation. In insulin-resistant T2DM and in obese patients, the PI-3K pathway, but not MAP-K, is resistant to insulin stimulation, resulting in a preferential activation of the MAP-K pathway [64]. MAP-K signaling in insulin-resistant patients is thought to have atherogenic and mitogenic properties that lead to atherosclerosis, while the anti-inflammatory, anti-atherogenic effects of physiologic levels of insulin, which depend on the PI-3K pathway, are lost. Compensatory hyperinsulinemia increases the production of endothelin and vascular adhesion molecules through a MAP-K dependent pathway. Consequently, cellular growth and migration is stimulated, together with production of prothrombotic and profibrotic factors, promoting atherosclerosis [65]. Furthermore, NO production is mediated via the PI-3K signal transduction pathway and is therefore impaired in insulin resistance states, promoting endothelial dysfunction and accelerated atherosclerosis.

Insulin resistance can promote atherosclerosis both by altering systemic processes like dyslipidemia and hypertension, as well as through impaired insulin signaling at the level of the cells of the vessel wall, including endothelial cells, VSMCs, and macrophages. These cells have insulin receptor-mediated signaling pathways that are repressed during hyperinsulinemia [66]. In endothelial cells, downregulation of insulin receptor levels through altered IRS/PI-3K pathways, results in reduced levels of eNOS activation and NO production, and increases in expression of the leukocyte adhesion molecule VCAM-1 involved in leukocyte entry into the vessel wall. Insulin can also activate the MAP-K dependent pathway which regulates secretion of the vasoconstrictor endothelin-1 (ET-1) from the endothelium. In VSMCs, insulin resistance may promote proliferation and migration of VSMCs into the subintimal space while also inducing VSMC apoptosis. The loss of VSMCs, an important cellular source of collagen and other matrix materials essential to fibrous cap formation, can lead to fibrous cap thinning and plaque necrosis. In monocyte-derived macrophages, insulin receptors are downregulated during obesity and hyperinsulinemia, and the impairment in insulin signaling promotes macrophage apoptosis in atherosclerotic lesions [67]. Elevated levels of saturated fatty acids, characteristic of insulin resistance state, can also amplify the apoptotic response in macrophages.

Relating preclinical studies of insulin signaling and its relationship to inflammation and atherosclerosis, to clinical issues regarding diabetes, can be challenging on several fronts. One of these involves the differences between acute insulin exposure, as can happen experimentally, and chronic hyperinsulinemia, which is encountered clinically in patients with insulin resistance. Insulin has been shown to have acute anti-inflammatory and vasodilatory effects. Chronic insulin stimulation, and other changes associated with chronic hyperinsulinemia, including insulin receptor changes, may have different effects than those seen in experimental models. The effects of endogenous insulin release and exogenous insulin administration may also have different effects. In the ORIGIN trial, subjects with recently diagnosed diabetes did not have any significant lowering of CV events with insulin therapy; at the same time, exogenous insulin had no adverse effects on promoting CV issues [68].

**Dyslipidemia, lipotoxicity, and lipid mediators of insulin resistance**

Dyslipidemia is a common finding in people with T2DM, who typically have increased plasma concentrations of TG and
low plasma levels of HDL-C. Diabetic patients have usually only mild elevation in plasma concentrations of LDL-C, but may have more atherogenic small, dense LDL. T2DM patients generally have more LDL particles than people without diabetes, and the increased LDL particles number may contribute to atherogenesis and pro-atherogenic effects. Increased LDL particles number may contribute to atherogenesis and CVD risk by increased entry into the arterial wall and an increased propensity for oxidation [69,70]. Meta-analyses clearly show that the reduction of LDL-C concentrations results in an important decrease in major coronary events in diabetic patients, at least if not more, than nondiabetic individuals [71,72]. Decreased HDL-C levels and a change in HDL composition, together with a decrease in the concentration of apolipoprotein A1, are observed in humans with T2DM. These changes impair the removal of excess cholesterol from macrophages and foam cells by HDL, which may be important in the early phase of the atherosclerotic plaque.

Lipotoxicity is a general term that refers to the deleterious effects of increased adiposity on metabolic responses [73]. Lipotoxicity is also used to refer to the effects of elevated FFA levels on the vasculature and other organs. Ectopic lipid accumulation, as is found in the visceral fat depot, skeletal muscle, and the liver, creates an imbalance between lipid uptake and oxidation, which is implicated in the pathogenesis of insulin resistance. Impaired ability of adipocytes to store additional TG, contributes to increased accumulation of lipids and their metabolites in other tissues. Excess of circulating FFA inhibits glucose transport and phosphorylation, with subsequent reduction of glucose oxidation and muscle glycogen synthesis, inducing insulin resistance in skeletal muscle [74]. The increase in liver fat also impairs the ability of insulin to regulate gluconeogenesis. Excess accumulation of lipids in the heart may promote cardiac lipotoxicity and development of myocardial dysfunction. Underlying molecular mechanisms invoked include increased lipid uptake or impaired mitochondrial oxidative function leading to accumulation of TGs and other toxic lipid compounds discussed later, which may cause myocyte loss through apoptosis and induction of prohypertrophic signaling [75].

Several toxic lipid metabolites including ceramide, DAG, and long-chain acyl-CoA are associated with lipotoxicity and contribute to the activation of PKC dependent signaling pathways and insulin resistance [76]. Accumulation of ceramides is associated with impaired PI-3K/Akt2 activation, subsequent increased expression of gluconeogenesis enzymes, and decreased insulin-mediated glycogen synthesis. Lipid excess with DAG activation of PKCs impairs insulin signaling and insulin receptor activation, contributing to insulin resistance in the muscle and liver [15]. Other kinases such as IκB kinase (IKK) complex and the c-Jun NH2-terminal kinase (JNK) are also activated by raised plasma FFA levels. FFA can reduce NO production through PKC-dependent stimulation of NADPH oxidase, increasing also the production of ROS [77]. Elevated FFA increases markers of endothelial activation such as ICAM-1, VCAM-1, and myeloperoxidase, and may facilitate endothelial damage by augmenting apoptosis. Consistent with this, hypertriglyceridemia is also associated with impaired vascular
reactivity. TG-rich lipoproteins enhance a pro-inflammatory phenotype in endothelial cells and increase TNFα expression and the adherence of monocytes and macrophages to endothelial cells [78]. Endogenous, physiologic hydrolysis of TG-rich lipoproteins generates endogenous PPARγ ligands that may limit inflammation and atherosclerosis [79].

Evidence from human studies suggests that increasing FFA levels by lipid infusion can inhibit insulin signaling and insulin-stimulated glucose disposal [80]. One challenge with many such studies is their reliance on heparin infusion, which displaces lipoprotein lipase (LPL) from the endothelial surface, allowing it to hydrolyze circulating TG-rich lipoproteins. Such methods do generate high levels of FFA but fundamentally disrupt the physiologic role of LPL on the endothelial surface.

In addition, FFA-induced impairment of endothelial function, has been reported as being reversed by an inhibitor of the renin-angiotensin system (RAS) and also by a PPAR gamma agonist [81,82].

**Metabolic inflammation and adiposity**

Obesity is associated with hypertension, dyslipidemia, and diabetes, all of which accelerate atherosclerosis. Increased adiposity alters lipoprotein metabolism, increases inflammation, and stimulates changes in adipocyte-generated cytokines, also known as adipokines (discussed further later), which help link obesity and CVD complications [83].

Increased circulating levels of inflammatory markers have been noted in obese individuals and in patients with the metabolic syndrome. Among nondiabetic subjects, elevated C-reactive protein (CRP) levels were related to measures of insulin resistance; the presence of inflammatory mediators may predict future occurrence of T2DM [84–86]. Furthermore, studies display a significant association between increased inflammatory markers, CV risk indicators, and CV events in T2DM [87–89]. Visceral fat accumulation is a predictor of CV events, suggesting that inflammation induced by adipose tissue may be a factor in the development of atherosclerosis and complications of the metabolic syndrome [90]. Obesity is characterized by low-grade chronic inflammation, which acts as a regulatory process that links multiple risk factors for atherosclerosis and its complications [91].

Inflammatory stimuli in obesity regulate secretion and activation of chemokines, adhesion molecule expression, and release of vasoactive mediators. Metabolic abnormalities such as dyslipidemia, hyperglycemia and insulin resistance augment the inflammatory response by multiple cellular pathways, as described earlier. The release of cytokines form a chronic low-grade inflammatory state (referred to by some as "metaflammation") generated by various cell types in response to an excessive energetic nutrient load which may be a key mechanism causing metabolic dysfunction and insulin resistance [92].

**Mechanisms of metabolic inflammation**

Inflammatory cytokines, such as TNFα, interleukins IL-6, IL-1β, CCL2 and others, are increased in adipose depots. In response to excess nutrients, cytokine activation in the setting of obesity is enhanced not only in the adipose tissue, but also in tissues like the liver, muscle and pancreas, which are also exposed to the inflammatory state induced in obesity. Adipose tissues display increased activation of intracellular kinases such as c-Jun N-terminal kinase (JNK), inhibitor of κ kinase (IKK) and protein kinase R (PKR), as well as the toll-like receptors (TLRs) of the immune system, stimulating downstream signaling pathways and inflammation. These inflammatory mediators may also contribute directly to insulin resistance as discussed. Activation of kinases results in serine phosphorylation of insulin receptor substrate 1 (IRS-1), inhibiting the insulin receptor signaling cascade [92]. Furthermore, these kinases can promote inflammation by activating transcription factors such as NFκB, which then increases expression of other inflammatory mediators.

Recent studies have drawn attention to the infiltration of inflammatory cells, including macrophages and T cells in visceral fat [93]. Cytokines in this setting can stimulate macrophage maturation in a "classical" (type M1) profile which is pro-inflammatory, in contrast to an "alternative" (type M2) phenotype which has anti-inflammatory effects, as noted earlier. Changes in the proportion of different T-cell types during obesity promote the phenotypic transition that occurs from M2 cells secreting anti-inflammatory molecules towards M1 macrophages releasing pro-inflammatory cytokines such as TNFα, IL-6, and IL-12 [94]. This line of investigation provides direct connects between inflammation, and inflammatory cells like macrophages, and the insulin resistance found during obesity and in T2DM.

Endoplasmic reticulum (ER) stress is also a novel pathway increasingly implicated in promoting inflammation in obesity and insulin resistance. The ER is the major site in the cell for protein folding and trafficking. Failure of the ER's normal adaptive capacity results in activation of the unfolded protein response (UPR), which intersects with many different inflammatory and stress signaling pathways [95]. ER stress may be further induced by hyperglycemia, increased fatty acids and hypoxia, and may initiate apoptotic pathways and production of oxidative stress, promoting inflammation and metabolic disease.

Gut microbiota influence metabolism in the obese state through bacteria and host interactions. Changes in the gut microbiota can affect insulin sensitivity and metabolic inflammatory state in adipose tissue and in peripheral organs [96].

**Adipokines**

Adipocytes secrete bioactive proteins called adipokines, which include various cytokines, chemokines, and hormones. These adipocyte mediators have been implicated in inflammation, especially in obesity in which adipose tissue is infiltrated by pro-inflammatory macrophages. Some of these adipokines are pro-inflammatory such as leptin, TNFα, and IL-6, while others have anti-inflammatory and protective effects against the formation of atherosclerotic lesions, such as adiponectin which is markedly reduced during obesity.
Adiponectin, which is produced more by subcutaneous fat than visceral fat, reportedly limits inflammation and atherosclerotic responses; adiponectin levels in humans are lower in obesity, insulin resistance, diabetes, and CVD [97]. Serum high-molecular-weight adiponectin levels are significantly lower in men with coronary artery disease, and adiponectin levels were shown to be inversely correlated to the progression of coronary artery calcium [98,99]. Adiponectin receptors activate the AMP kinase (AMPK) pathway and the peroxisome proliferator-activated receptor alpha (PPARα) pathway in the liver, increasing insulin sensitivity and decrease inflammation. Adiponectin reportedly suppress VSMCs proliferation and transformation of macrophages to foam cells, with reduction of lipid accumulation in foam cells. Adiponectin may prevent proliferation and transmigration of fibroblasts in the vascular adventitia. Expression of adiponectin is under the control of PPARγ.

Leptin is a central nervous system hormone involved in appetite regulation. Plasma leptin levels are increased in obesity. Leptin may also produce pro-inflammatory effects, enhancing cellular immune responses, proliferation of macrophages, activation of cytokines, and increases in ROS production. Leptin may also have procoagulant effects. Taking these factors together, leptin may be involved in modulating atherogenesis. A study in T2DM demonstrated that increased leptin levels are associated with coronary atherosclerosis, independent of insulin resistance [100]. Not unlike insulin, leptin resistance has also been suggested as a clinical syndrome. Another important adipokine is resistin, which is expressed in macrophages of atherosclerotic lesions. Resistin induces the expression of adhesion molecules, promoting the adhesion of monocytes to the vascular wall. It is associated with insulin resistance and coronary artery disease [101,102].

Additional obesity-induced inflammatory mediators in adipose tissue include TNFα, which stimulates lipolysis and inflammation and is associated with insulin resistance. Circulating IL-6 is increased in obese people, regulating immune response and inflammation, and contributing to insulin resistance in the liver and muscle. Proteins that recruit inflammatory cells to specific sites are known as chemoattractant cytokines, or more simply chemokines. Monocyte chemoattractant protein-1 (MCP-1), an important chemokine released by endothelial cells and adipocytes, helps recruit monocytes and other cells expressing the MCP-1 receptor to adipose tissue and elsewhere. Adipocytes also produce the procoagulant and CV risk factor PAI-1, thus connecting adiposity to hypercoagulability [103]. Levels of angiotensinogen, a central player in the renin-angiotensin system, also correlates with accumulation of adipose tissue, and has been reported to stimulate lipogenesis and vascular inflammation, connecting these responses to hypertension.

Increased visceral fat results in dysregulation of the different adipokines, contributing to the development of metabolic syndrome, and to the generation of pro-inflammatory, pro-thrombotic, and ultimately pro-atherosclerotic components. Adipose tissue distribution is important for atherosclerotic

---

**Figure 70.4** Adiposity and vascular responses. Excess adiposity exerts multiple cellular effects that can promote inflammation and atherosclerosis. Obesity, especially increased visceral adiposity and other sites of increased lipid storage in abnormal locations (liver, muscle), is associated with excess FFA and specific lipid metabolites, alteration in hormones and inflammatory mediators like cytokines. Together these factors are thought to increase insulin resistance and promote inflammation via multiple signaling pathways and inflammatory mediators. TNFα, tumor necrosis factor alpha; IL-6, interleukin 6; PAI-1, plasminogen activator inhibitor 1; MCP-1, monocyte chemoattractant protein-1; FFA, free fatty acids; JNK, Jun N-terminal kinase; IKKβ; inhibitor of k kinase beta.
risk. Visceral, truncal fat is associated with more vascular and metabolic risk than peripheral fat, possibly due to the pattern of adipokines and cytokine release from visceral fat, and its access to the portal circulation [83].

Overall, metabolic inflammation is characterized by a low-grade chronic inflammation, induced during obesity in response to excess nutrients, by release of inflammatory cytokines, adipokines, and macrophages [92]. Impaired signaling pathways and triggers such as ER stress, hypoxia, ROS generation, and altered gut microbia, further maintain the chronic inflammatory state, contributing to insulin resistance and metabolic dysfunction (Figure 70.4).

**Conclusion**

The incidence of obesity and diabetes is rising worldwide and will impact increasingly on the nature and care of CVD. Diabetes elevates the risk for CVD complications, particularly atherosclerotic vascular disease. The cluster of abnormalities typically associated with diabetes and often prediabetes known as the metabolic syndrome or insulin resistance syndrome incorporates various CVD risk factors, and predicts future diabetes and coronary artery disease. Inflammation, including adipose inflammation, and excess adiposity, are emerging as a key mechanism for the metabolic derangement found in T2DM, with considerable overlap with the changes seen in the arterial wall with diabetic atherosclerosis. Hyperglycemia, insulin resistance and lipoprotein abnormalities, are all connected through complex pathophysiologic interactions, further promoting accelerated atherosclerosis and CVD in diabetes. New insight has been generated into molecular mechanisms involved in these processes. These mediators provide new opportunities for better predictive biomarkers and therapeutic targets.

**References**


Atherogenesis, coronary heart disease and insulin resistance syndrome in diabetes

Endothelial function and metabolic syndrome

Xinpu Chen and Sridevi Devaraj
Texas Children's Hospital, Baylor College of Medicine and Texas Children's Microbiome Center, Houston, TX, USA

Key points
- Metabolic syndrome is highly prevalent and predisposes to increased risk of diabetes and cardiovascular disease.
- Metabolic syndrome is a pro-inflammatory state.
- Endothelial dysfunction is a key early event in atherogenesis and is pronounced in patients with metabolic syndrome.
- The measurement of endothelial function in vivo is far from optimum and newer biomarkers are needed.
- Two biomarkers of endothelial function that hold promise are the evaluation of endothelial microparticles and endothelial progenitor cells.

Introduction: the metabolic syndrome

The metabolic syndrome (MetS) comprises a cluster of cardiometabolic risk factors, with insulin resistance (IR) and central adiposity being the main features. The National Cholesterol Education Program Adult Treatment Panel III (NCEP-ATP III) defined the MetS as the presence of any three features of the five outlined (central obesity, dyslipidemia (i.e., high triglycerides, low HDL), hypertension, and impaired fasting glucose) [1–3]. Approximately 35% of US adults have the MetS and this is a global health issue. Also, the prevalence of MetS increases with age and is more in certain ethnic groups. MetS confers a two- to fourfold increased risk for cardiovascular disease (CVD) and fivefold increased risk of diabetes [1–3]. Several lines of evidence have reported that MetS is a proinflammatory state.

Endothelial dysfunction and MetS

The endothelium plays an important role in maintaining vascular health and integrity and secretes endothelium-derived relaxing and endothelium-derived contracting factors, such as nitric oxide, which plays a pivotal role in the maintenance of vascular tone and reactivity. Diminished production or availability of nitric oxide and/or imbalance in the ratio of endothelium-derived relaxing and contracting factors results in endothelial dysfunction. Endothelial dysfunction is commonly associated with the development and progression of a wide range of cardiovascular diseases. A key early event in atherosclerosis is endothelial cell dysfunction, which is precipitated by several noxious insults including obesity, hypertension, dyslipidemia, hyperglycemia (all features of MetS), as well as other insults including smoking [4–6]. Many investigators have reported endothelial dysfunction in patients with MetS. One of the most widely used measurements of endothelial function is flow-mediated dilatation (FMD), which is a nitric oxide-dependent, endothelium-dependent vasodilatation [7].

In one of the first reports on endothelial dysfunction and its association with MetS, Esposito et al. [8] showed that compared with 60 control subjects matched for age and sex, patients with the metabolic syndrome had decreased endothelial function. In the Framingham Offspring participants, Hamburg et al. [9] demonstrated that MetS was associated with decreased FMD. Furthermore, as the number of MetS components increased, there was worsening of endothelial dysfunction. Lteif et al. [10], using leg blood flow measurements, showed that patients with MetS had worse endothelial function. Also, in the Prospective Study of the Vasculature in Uppsala Seniors (PIVUS) study [11], using different techniques to assess vasodilatation in conduit and resistance arteries in MetS, the authors showed decreased FMD in patients with MetS. In the Northern Manhattan study (NOMA), Suzuki et al. [12] reported that MetS was associated with decreased FMD and increased cardiovascular disease over an 81-month follow-up period. All of the above reports clearly document that MetS patients have impaired endothelial function. Since endothelial vasodilator function is important in maintaining vascular integrity, the loss of endothelial...
function would be expected to result in increased cardiovascular burden. This has major implications with regards to subsequent CVD.

**Endothelial microparticles (EMP)**

Inflammation also begets increased oxidative stress and apoptosis which results in cellular release of microparticles (MP). MPs are small membrane fragments that are shed from various cell types and range between 0.1 and 1.0 μm [13–17]. The release of MPs is regulated by different stimuli such as shear stress, physiologic agonists, proapoptotic stimulation, and damage. MPs are membrane vesicles with procoagulant and proinflammatory properties. MPs are present in blood from healthy individuals and elevated in patients under pathologic states, such as sepsis, various clinical situations associated with thrombosis, and in diabetic patients, strengthening the notion that MPs may play a role in these diseases [13–16]. Indeed, MPs can be considered as vectors of biologic messages, such as induction of endothelial and vascular dysfunctions.

The endothelium is one of the primary targets of circulating MPs. Under normal conditions, MPs contribute to the regulation of endothelial cell functions, including coagulation and inflammation. MPs derived from endothelium are referred to as endothelial microparticles and appear to contribute to a procoagulant and proinflammatory phenotype that leads to endothelial dysfunction [18].

**Endothelial microparticles and MetS**

In MetS, few groups have studied EMPs [19,20]. Tissue factor-positive MPs are associated with components of the MetS in patients with well-regulated type 2 diabetes. Recently, Arteaga et al. [19] reported endothelial cell MP release, and increased binding of MPs from endothelial cells and platelets to leukocytes in patients with the MetS. Chironi et al. [20] showed in patients with MetS, that the leukocyte-derived MP level is higher than in those free of such syndrome and increased gradually in parallel with the number of components of MetS.

In a large study of 88 subjects with MetS, Helal et al. [21] showed significantly increased EMPs in these patients compared to matched controls and they correlated significantly with the features of MetS, namely, waist circumference, blood pressure, and dyslipidemia.

In conclusion, data from these studies demonstrates that elevation of circulating EMPs is associated with the cluster of abnormalities characteristic of the MetS, and could be an early predictor of metabolic risk. Although MP are known to play a pathogenic role in more advanced stages of atherosclerosis, they could also play a detrimental role in conditions such as the MetS. Further studies are required to determine the validity of using MP as biomarkers for progressive metabolic abnormalities in MetS.

**Endothelial progenitor cells (EPC)**

In recent years, there has been much interest in a subtype of progenitor cells, isolated from bone marrow, umbilical vessels, and peripheral blood of adults which have the capacity to circulate, proliferate, and differentiate into mature endothelial cells, termed endothelial progenitor cells (EPCs) [22–25]. EPCs circulate in the blood and appear to home preferentially to sites of vascular or tissue injury, contributing significantly to both re-endothelialization and neoangiogenesis. There are, however, many controversies with regard to the exact definition and characterization of human EPCs. The most often used characterization of EPCs is the assessment of surface markers such as CD34 and vascular endothelial growth factor receptor-2, VEGFR-2 (KDR). Indeed, the assessment of CD34+KDR+ combination is the only EPC phenotype that has been demonstrated repeatedly and convincingly to be an independent predictor of cardiovascular outcomes.

**EPC and cardiovascular events**

Decreased EPC number and functionality is associated with increased cardiovascular events [26–30]. In a 10-month follow-up study, Schmidt-Lucke et al. [26] showed that the level of CD34+KDR+ cells independently predicted cardiovascular events and progression of atherosclerosis in a mixed population of healthy subjects and cardiovascular patients. In a larger study, Werner et al. [27] have reported that CD34+KDR+ cell count predicted cardiovascular events and cardiovascular death during a 12-month follow-up in 519 patients with coronary artery disease (CAD). Also, in a subset, colony-forming units (CFUs) predicted cardiovascular events. Furthermore, Hill et al. [28] reported a strong correlation between the number of circulating endothelial progenitor cells (measured as CFUs) and the subjects’ combined Framingham risk factor score. Also, the measurement of flow-mediated brachial-artery reactivity revealed a significant relation between endothelial function and the number of progenitor cells. Indeed, levels of circulating EPCs were a better predictor of vascular reactivity than was the presence or absence of conventional risk factors [29]. Fadini et al. showed that a low CD34 count, a measure of progenitor cells, in addition to MetS was associated with increased cardiovascular events (CVEs). Fadini’s group have also shown an association between EPC reduction and increased carotid intima media thickness (c-IMT), as a marker of early atherosclerotic remodeling in healthy subjects [30,31].

EPC functionality is determined by its capacity to form colonies, migration, and ability to form tubules. In addition to flow cytometric quantitation of CD34/KDR predicting CVE, functional assays such as CFU and EPC migration have been shown to correlate significantly with CAD risk factors, severity, and events [29].
Thus, the measurement of EPCs may be a surrogate biologic marker for vascular function and cumulative cardiovascular risk, suggesting further that endothelial injury in the absence of sufficient circulating progenitor cells may unfavorably affect the progression of CVD. Furthermore, decrease in circulating EPCs contributes to impaired angiogenesis as well as progression of atherosclerosis and patients at risk for CAD have a decreased number of circulating EPCs with impaired activity. Thus, it seems important that both the number and functional activity of EPCs should be investigated. The individual components of the MetS are associated with impairment of EPC number and function.

**EPC and MetS**

With regards to the MetS, there is limited data on EPC number and functionality. In the study by Westerweel et al. [32] they show that circulating CD34+KDR+ EPC levels were reduced by nearly 40% in obese men with MetS compared to nonobese men. Although this was a small study that included 19 patients with MetS, it is important to emphasize that in this study, they excluded patients with overt clinical CVD or diabetes. They did not study EPC functionality. In the study by Jialal et al. [33], they reported on EPC number and functionality in a larger sample size of subjects with MetS (n = 46, without the confounding factor of diabetes or cardiovascular disease) of which 77% were female and matched controls (n = 31). In accord with the study in obese males, they showed a significant decrease in EPC number, also defined by CD34/KDR dual positivity. Furthermore, these investigators also looked at functionality of EPCs such as CFUs, migration, and tubule formation. In addition to the reduction in numbers, they showed that there was significant impaired clonogenic capacity and also an impaired capacity to incorporate into tubule structures. Whilst there was a decrease in migration of the EPCs in MetS this did not attain significance. It needs to be emphasized that none of the subjects were on medications that affect EPCs suggesting that the defect in EPCs manifests early in nascent MetS prior to the development of diabetes or CVD. Fadini et al. [30] have reported in a study decreased circulating EPCs and progenitor cells in diabetic patients with peripheral vascular disease. In this paper, they did a subgroup analysis of MetS patients versus non-MetS patients. However, not much detail is provided with respect to coexistent diseases and morbidity such as diabetes and peripheral vascular disease or concomitant medications in these two subgroups. In a subsequent report by Fadini et al. [31], they showed that in patients with MetS, there was a decrease in progenitor cells (CD34+ cells). It appears that many of these patients also could have diabetes, and be on medications such as statins, angiotensin-converting enzyme inhibitors (ACE-I), angiotensin receptor blockers (ARBs), and anti diabetic therapy such as pioglitazone, which could have influenced the data. Previously, Satoh et al. have reported an increased EPC number in CAD patients with MetS and without MetS [34]. They did not compare patients with MetS with controls and their sample sizes were small (n = 15 for acute myocardial infarction and n = 16 for patients with stable angina pectoris and MetS, respectively). Interestingly, they also showed increased oxidative DNA damage, decreased telomerase activity, and decreased telomere length, a marker of increased senescence in EPCs of CAD patients with MetS than the CAD patients without MetS. This suggests that the increase in EPC with CVD was a dysfunctional population since EPCs are generally well endowed with antioxidant defenses. Other functional measures of EPC activity such as tube formation or CFUs or adhesion were not investigated in this study. Thus, this needs to be investigated further. Recently, La Vignera et al. [35] reported increased EPCs in patients with arterial erectile dysfunction and MetS compared to controls. Furthermore, the increased EPC in their subjects correlated with EMPs and IMT suggesting that this is a dysfunctional population. However, they did not study EPC functionality.

MetS is a proinflammatory state. In the study by Jialal et al. [33], they showed significant correlation of CRP levels in MetS with decreased EPC number and functionality, pointing to the role of inflammation in this process.

There is limited data with regards to mobilizing factors in patients with MetS. Egan et al. [36] have reported the profound reduction in EPCs due to impaired mobilization from bone marrow because of the lower expression of CXCR chemokine receptor 4 (CXCR4+)/CD34+ cells in diabetics versus controls. Importantly, CXCR4, CD117, and KDR are defined as the mobilizing receptors for progenitor cells (PCs). Jialal et al. recently showed in subjects with MetS (n = 36) compared to age- and gender-matched controls (n = 38) that there was a significant reduction of 83% in granulocyte colony-stimulating factor levels in patients with MetS [37]. Also, there were decreases in SCF and SCF soluble receptor levels. However, there was no significant difference in stromal cell-derived factor-1 levels, and paradoxically, vascular endothelial growth factor levels were increased, consistent with VEGF resistance, which has been reported previously with insulin-resistant states such as diabetes and MetS.

In conclusion, EPC number and functionality could serve as an additional novel cellular biomarker of endothelial integrity and impaired neoangiogenesis in patients with MetS who clearly have manifest endothelial dysfunction [38]. Prospective studies should demonstrate that they predict CVD. Thus, EPC number and/or functionality could emerge as a novel cellular biomarker of CVD risk and could better inform clinicians about potential pharmacotherapy for patients with MetS.

**Conclusion**

Thus, this review of the literature indicates that endothelial dysfunction is pronounced in MetS and while there is much variability in the study of endothelial dysfunction by...
flow-mediated dilatation, several new cellular markers such as endothelial microparticles and endothelial progenitor cells may emerge as novel biomarkers of endothelial function, which can also be used to test management/pharmacotherapy of MetS, as it relates to vascular dysfunction.

**References**


CHAPTER 72

Hemostatic abnormalities in diabetes mellitus

R.A. Ajjan and Peter J. Grant

Division of Cardiovascular and Diabetes Research, University of Leeds/Leeds Teaching Hospitals Trust, Leeds, UK

Key points

• Diabetes is a chronic progressive metabolic state characterized by the presence of inflammatory thrombotic changes that promote vascular damage.

• In insulin-resistant states, suppression of fibrinolysis due to elevated levels of the fibrinolytic inhibitor, plasminogen activator inhibitor-1, is an invariable finding. Elevated levels of fibrinogen and other clotting factors are also seen.

• Hyperglycemia further inhibits lysis by inducing qualitative changes in coagulation and fibrinolytic factors altering protein activity and generating a fibrin clot resistant to lysis.

• The platelet responds to the complex metabolic milieu associated with advanced diabetes by developing enhanced activation, increased aggregatory and adhesive properties and interacting with monocyte/macrophages to promote an inflammatory thrombotic environment.

• Disturbances in circadian rhythm may provide an important link between obesity, inflammation, and abnormalities in thrombosis.

• Interventions that ameliorate insulin resistance and hyperglycemia have potentially beneficial effects on the abnormalities in coagulation/fibrinolysis, and platelet function seen in diabetes.

Introduction

Hemostatic mechanisms in man have a primary physiologic function to arrest bleeding in the context of trauma and vascular damage, with a secondary role in vascular repair. Interactions between the fluid and cellular phases of thrombotic pathways (coagulation/fibrinolysis and platelets) and the inflammatory system (complement, cytokines, macrophages) are important in integrating thrombotic and anti-infective processes in these situations. Abnormalities in hemostasis are associated with a bewildering array of bleeding disorders, primarily linked to deficiencies in clotting factors and platelets, and a range of thrombotic diseases affecting both the venous and arterial systems.

Plasma levels and/or activity of components of coagulation/fibrinolytic pathways and platelets are regulated by a wide variety of genetic and environmental influences, the latter including the direct effect of post-translational modifications and the indirect influence of other factors such as dyslipidemia, smoking, insulin resistance, and obesity. Diabetes mellitus is itself associated with a plethora of metabolic changes many of which interact with components of the hemostatic system to increase thrombotic risk which clinically manifests with increased rates of myocardial infarction, thrombotic stroke, peripheral vascular disease, and venous thrombosis.

Occlusive thrombotic vascular disease is a major cause of morbidity and mortality in subjects with diabetes and the majority of patients with diabetes ultimately die from vascular causes. Arterial disease is characterized by the early development of endothelial dysfunction and fatty streaks followed by the development of arterial plaques, and finally occlusive thrombus formation on a ruptured, unstable plaque. Diabetes affects all aspects of these processes and clinical studies indicate that coronary artery plaques from subjects with diabetes have increased plaque thrombus and monocyte/macrophage infiltration compared to nondiabetic controls. This together with more extensive disease affecting both the proximal and distal coronary vasculature describes a situation in which the circulation supplying the heart has more lesions, with a greater propensity to rupture and to produce more thrombus. Arterial thrombus is characterized by the development of a platelet-rich fibrin mesh, whilst venous thrombosis is characterized by a fibrin-rich, platelet-poor thrombus. Type 2 diabetes (T2DM) is associated with increased platelet activation and a range of abnormalities in coagulation and fibrinolysis related to the metabolic abnormalities associated with insulin resistance and hyperglycemia [1]. These prothrombotic changes contribute to
the increased prevalence and severity of acute coronary syndromes and other arterial disorders. Increased platelet reactivity in particular has been reported to prospectively predict risk of major adverse cardiovascular events in T2DM patients with stable coronary artery disease [2]. Most evidence seems to indicate that thrombotic disorders appear with the development of insulin resistance and in the presence of complications such as microalbuminuria and chronic kidney disease [3,4]. Glycemia has an additional effect on many of these processes which tend to deteriorate as the chronic nature of diabetes unfolds. Clinical studies suggest that as a consequence un-complicated T1DM has relatively minor alterations in thrombotic profile, whilst nondiabetic insulin-resistant relatives of subjects with diabetes have clustering of inflammatory thrombotic risk prior to the appearance of frank hyperglycemia [5,6] and in both groups further changes occur as the disorder progresses.

The recognition that myocardial infarction usually results from thrombus formation on a ruptured plaque has led to a revolution in therapeutic approaches which has improved primary and secondary prevention of cardiovascular disease as well as the management of acute coronary syndromes. Amongst these, the development of increasingly sophisticated inhibitors of platelet activation, direct thrombin inhibitors and heparin-like molecules have transformed care of both diabetic and nondiabetic subjects with coronary artery disease. In this chapter we will describe the mechanisms that underpin abnormalities in platelet function and the fluid phases of coagulation/fibrinolysis in subjects with diabetes and the way in which these changes relate to cardiovascular disease.

**Mechanisms of thrombosis**

**The hemostatic system**

The hemostatic system consists of a fluid phase of activators and inhibitors of coagulation and fibrinolysis which regulate the formation and breakdown of fibrin and a cellular, platelet phase which interacts with sites of vascular damage and fibrin to release a range of procoagulant and inflammatory mediators. The interaction between the fluid and cellular phase of coagulation is crucial to maintaining a normal hemostatic environment and thrombin is the pivotal enzyme with a key role in both fibrin formation and platelet activation. Thrombin is generated by the cleavage of prothrombin by a Factor Xase complex which occurs as the result of interactions between tissue factor, activated Factor VII and Factor X secondary to vascular damage. Thrombin, whilst having major procoagulant and pro-inflammatory effects, can express an anticoagulant effect when thrombin binds to the cell-associated receptor thrombomodulin, which modulates clot resistance to fibrinolysis.

**Fibrinogen**

The insoluble fibrin network, forming the backbone of the blood clot, is derived from the soluble fibrinogen molecule. Fibrinogen is a large protein produced by the liver which consists of two sets of α, β, and γ chains linked by disulfide bonds. Thrombin cleaves fibrinogen by cutting small fibrinopeptides from each of the fibrinogen α chains allowing the α chains to open up and interact with other cleaved fibrinogen molecules leading to the formation of double-stranded fibrils which branch out to create a complex fibrin network [7]. Cleavage of fibrinopeptide B allows lateral aggregation of the developing fibrin structure. The fibrinogen molecule has another important role in thrombotic processes, acting as a ligand for the activated platelet GpIIb/IIIa receptor. Binding to this receptor facilitates platelet/platelet interactions and platelet aggregation.

**Factor XIII activation and fibrin cross-linking**

Coagulation FXIII is a transglutaminase which circulates in plasma in a heterodimeric structure that consists of two A- and two B-subunits. Thrombin activates Factor XIII by cleaving a 37 amino acid peptide from the A subunit which promotes separation of the A- and B- subunits and permits exposure of the active site on FXIII A. Activated Factor XIII A covalently cross-links fibrin fibrils and various antifibrinolytic proteins to the clot which creates a fibrin network with altered mechanical properties and increased resistance to fibrinolytic activity [8].

**Fibrinolysis**

Analogous to thrombin, plasmin is the pivotal enzyme in the fibrinolytic cascade. Plasmin is generated by the cleavage of plasminogen by tissue plasminogen activator (tPA) and this reaction occurs 1000-fold faster in the presence of fibrin [9]. A lysine binding site on plasmin binds plasmin to fibrin which facilitates fibrin breakdown and also protects plasmin from local inhibition by antiplasmin. Plasmin cleaves arginine and lysine sites on a range of molecules and its activity is tightly controlled by antiplasmin to prevent systemic proteolysis. Cleavage of fibrin by plasmin leads to the generation of fibrin degradation products. Fibrin degradation products in turn can be measured in plasma, and one of which, D-dimer, is used as an indicator of the presence of venous thrombotic disease [10]. In addition to antiplasmin, other inhibitors of this pathway include plasminogen activator inhibitor-1 (PAI-1) and thrombin-activatable fibrinolysis inhibitor (TAFI). PAI-1 is the fast-acting inhibitor of tPA which binds to and inhibits tPA activity. PAI-1 is produced by endothelial cells, platelets liver and adipose tissue, and circulates in plasma in excess over tPA and is also found in fairly high concentrations in thrombus [1]. TAFI is found in large quantities in platelets and plasma and is activated by thrombin, a cleavage event that is much enhanced when thrombin is bound to thrombomodulin. Activated TAFI cleaves the N-terminal lysine residues from degrading fibrin to prevent binding of plasminogen to fibrin, consequently resulting in inhibition of plasmin generation and clot lysis.

**Platelet activation**

Damage to the vascular wall leads to two key events in platelet-associated clot formation: (1) receptor-mediated
Mechanisms of thrombosis in diabetes

Fibrinolysis

The fibrinolytic process is influenced both by the structure of the fibrin network, as compact clots are more difficult to lyse, and abnormalities in fibrinolytic proteins.

Plasminogen activator inhibitor-1 (PAI-1)

The major consistent hemostatic abnormality observed in insulin-resistant T2DM is marked suppression of fibrinolysis associated with increased levels of both PAI-1 and tPA [14]. Studies of euglycemic first-degree relatives of subjects with T2DM indicate that such individuals tend to be insulin resistant and have raised tPA and PAI-1 before a diagnosis of diabetes is made, indicating that glycemia per se is not the main factor responsible for altered protein levels. The Insulin Resistance Atherosclerosis Study reported that early increases in PAI-1 in subjects with the metabolic syndrome predicted conversion to T2DM and suggested that high PAI-1 may be a marker for risk of developing the metabolic syndrome. In the Leeds Family Study of clinically healthy individuals, clustering of conventional risk factors was associated with progressive suppression of global tests of fibrinolysis that was related to the number of risk factors an individual held [15]. These findings in a randomly selected cohort of families support the view that increases in PAI-1 occur early in the course of the development of cardiovascular disease. The PAI-1 gene has a 4G/5G polymorphism 675bp from the start site and the 4G allele has been associated with both higher PAI-1 levels and increased risk of acute coronary syndrome [16]. Additionally there are indications that interactions between the 4G/5G genotype and triglyceride generate a nuclear complex that regulates PAI-1 expression providing a mechanism for increasing cardiovascular risk [17]. In nondiabetic populations the link to cardiovascular risk is provided by prospective studies indicating that elevated PAI-1 predicts the development of platelet adherence and aggregation, and (2) thrombin-mediated platelet activation. Adherence to the subendothelial matrix is facilitated by a range of glycoprotein receptors (GP Ib/IX, GPVI, and GPla) which interact with von Willebrand factor to promote platelet adhesion. This interaction leads to activation of platelet GPIIb/IIIa which binds fibrinogen and promotes platelet aggregation. Thrombin is the most potent platelet activator which exerts its effects through binding to protease-activated receptor 1 (PAR-1) on the platelet surface. This leads to a cascade of signaling processes culminating in the release of a range of inflammatory and thrombotic mediators that further promote clot formation. In addition to thrombin, a range of other mediators including ADP, collagen and thromboxane can activate the platelet through a receptor binding event. These receptors provide some of the novel targets for therapeutic approaches discussed later and described in a number of excellent reviews [11,12]. A summary of fibrin clot formation and lysis following vascular damage is provided in Figure 72.1.

Summary

An important characteristic of thrombotic processes is the closely regulated interactions between soluble circulating proteins involved in coagulation/fibrinolysis and the cellular component of coagulation. In addition to platelets, thrombosis involves binding events on endothelium, the exposed subendothelial layer, macrophages and leukocytes with the balance between thrombosis and clot lysis and the localization of clot formation depending on these interactions. Emerging evidence demonstrates the importance of thrombotic inflammatory interactions, both at the cellular level where platelet/macrophage binding initiates the release of a range of soluble procoagulant and inflammatory molecules and in the fluid phase where, for example, complement C3 binds fibrin to inhibit fibrinolysis [13]. As these events cycle towards fibrin formation and fibrin/platelet interactions, further levels of control are exerted by the interaction of activators and inhibitors of lysis on fibrin itself. All these levels of control direct and limit thrombus formation and fibrinolysis to the site of damage to prevent systemic thrombus formation and proteolysis.

Figure 72.1 Formation of the fibrin clot and fibrinolysis. Injury to the vessel wall results in platelet aggregation and activation of the coagulation pathway culminating in thrombin production. Thrombin cleaves the fibrinogen molecule transforming the soluble protein into an insoluble network of fibrin fibers. Thrombin-activated factor (F)XIII further stabilizes the fibrin clot by cross-linking fibrinogen chains and enhances resistance to fibrinolysis by cross-linking various antifibrinolytic proteins including plasminogen inhibitor and thrombin-activatable fibrinolysis inhibitor. Plasmin, generated from plasminogen by tissue plasminogen activator (tPA), is the main protein responsible for clot lysis, resulting in the generation of fibrin degradation products, one of which is D-dimer. The fibrinolytic process is closely controlled by plasminogen activator inhibitor (PAI)-1 and increased levels of this protein can result in a pathologic hypofibrinolytic environment, commonly seen in insulin-resistant states including type 2 diabetes.
myocardial infarction, and gene association studies which relate the 4G allele in the PAI-1 gene to history of cardiovascular disease. However, there are considerable inconsistencies in the literature regarding these associations with both positive and negative studies. Authors have tried to explain this by various means including a requirement for PAI-1 to have the presence of additional risk factors (the ECTIM study) and a potential effect of the PAI-1 gene on cholesterol metabolism [18]. In a small study of 160 subjects with T2DM, the PAI-1 4G allele was significantly more common in the 38 subjects with clinical evidence of coronary artery disease. In modern terms, this is a tiny case control study and the results may represent skewed data. In nondiabetic patients with acute stroke no association was found between PAI-1 genotype and thrombotic stroke and in 40T2DM subjects with acute stroke compared to 80 non-stroke T2DM measures of circulating parameters of fibrinolysis did not suggest an association between suppression of lysis and cerebrovascular disease.

The range of cell lines that produce PAI-1 has produced a variety of results that have largely not been reproduced in clinical studies. In vitro PAI-1 synthesis and secretion can be induced by insulin, insulin like growth factor-1, triglyceride, and glucose. In Hep G2 cells, insulin is reported in some, but not all, studies, to increase PAI-1 expression. Synergistic responses have been reported with insulin and a combination of triglyceride and free fatty acids in Hep G2 cells, a finding reproduced in human subjects [19,20]. One of the most interesting developments in this field has, however, related to the potential role of obesity and the adipocyte in regulating PAI-1 expression. Human adipocytes in vitro express PAI-1 in response to insulin, angiotensin II, fatty acids, and TNFα. The addition of the TNFα inhibitor pentoxifylline to adipocytes decreases basal PAI-1 expression to suggest that autocrine regulation of PAI-1 by adipocyte TNFα may be important, linking adipocyte inflammatory responses to thrombotic risk [21]. This hypothesis was partially tested in vivo using open heart surgery as a model of a severe inflammatory response and in this study 27-fold increases in PAI-1 gene expression were seen in omental fat cells and threefold increases in subcutaneous fat cells. This increase in gene expression was associated with an increase in circulating PAI-1 levels [22]. Further sophistication of this model has been generated by the role of the macrophage in adipocyte responses. It is now established that in obesity the fat cell mass is infiltrated by macrophages that are activated by overspill of free fatty acids from the adipocyte and which in response release a range of cytokines, including TNFα to stimulate adipokine expression, including PAI-1 [23]. These findings provide further support for the emerging view that inflammatory thrombotic interactions at the cellular and protein level underpin many of the critical processes that generate atherothrombotic risk in diabetes.

**Thrombin-activated fibrinolysis inhibitor (TAFI)**

TAFI levels seem to be unaffected by insulin resistance or hyperglycemia, although there are indications that levels are increased in T2DM with microalbuminuria [24].

**Plasminogen activation**

We have recently observed that inhibition of fibrinolysis in poorly controlled diabetes is partially reversed by relatively minor improvements in glycemia (around 1% decrease in HbA1c), an effect independent of PAI-1 [25]. We have demonstrated that deranged lysis in diabetes is partly related to impaired plasmin activation secondary to increased protein glycation and that this can be ameliorated by improving glycemic control, consequently resulting in improved plasmin activity and fibrinolysis [26]. These observations suggest that management of lifestyle, weight loss, insulin resistance, and glycemia all benefit the hypofibrinolytic state observed in subjects with diabetes.

**Coagulation**

The coagulation process involves a number of protein interactions, which were briefly discussed earlier, and a more detailed description of this process can be found elsewhere [27]. The role of key coagulation proteins in atherothrombosis is discussed later.

**Factor VII**

Factor VII is a vitamin K-dependent serine protease synthesized in the liver which is activated by tissue factor released from damaged tissue and which has a key role in activation of the coagulation cascade and conversion of prothrombin to thrombin. In a manner similar to that seen with PAI-1, FVII levels show an association with insulin resistance and T2DM, with increased levels described in healthy individuals with the metabolic syndrome, first-degree relatives of T2DM patients and in type 2 diabetes itself [1]. Regulation of FVII appears to take place through the effects of triglyceride-rich particles and in clinical studies levels of FVII correlate with plasma triglycerides. It has been postulated that deficient postprandial catabolism of triglyceride may prolong the half-life of FVII, thereby increasing levels, and/or that triglyceride binds and activates FVII [1]. The role of FVII as an independent predictor of cardiovascular events is open to question although the 16-year follow-up of the Northwick Park Heart Study reported a significant association between elevated FVII and fatal cardiac events.

**Fibrinogen**

Plasma levels of fibrinogen influence clot formation (see later), blood flow, and platelet aggregation and elevated levels are a consistent cardiovascular risk factor in both diabetic and nondiabetic populations [28]. Although the relationship between fibrinogen and features of the insulin resistance syndrome is weaker than for PAI-1 and FVII, consistent associations exist with insulin levels, BMI, and reduced HDL [1]. Fibrinogen levels are elevated in healthy first-degree relatives of patients with T2DM and in T2DM in both Caucasian and Asian subjects [29]. A positive correlation has been reported between fibrinogen and glycemic control, although improvements in glycemia do not consistently reduce fibrinogen levels. Glycation of fibrinogen occurs on lysine residues in the β- and γ-chains of fibrinogen, an effect which appears to alter fibrin structure/function (see fibrin,
Later) but not acetylation by aspirin [30]. Metformin and thiazolidinedione therapy in T2DM leads to a fall in fibrinogen, in support of insulin resistance regulating fibrinogen expression [26].

**FVIII/von Willebrand factor (vWF)**

Synthesized by endothelial cells and megakaryocytes, vWF has an important role in protecting and therefore increasing the half-life of Factor VIII in the circulation. vWF acts as a ligand for the Gp-Ib platelet receptor and acts to tether platelets to the collagen-rich subendothelial layer exposed by vascular damage. Analogous to the platelet aggregation effects of fibrinogen, vWF enhances platelet adhesion and these two proteins have a crucial role in the formation of a platelet-rich clot characteristic of arterial disease. Coagulation Factor VIII is a component of the contact activation pathway, deficiency of which leads to hemophilia. Evidence indicates that both FVIII and vWF associate with features of the metabolic syndrome [31] and may have a role in arterial disease, although as with other coagulation factors, independent associations are open to question.

**Factor XII**

Factor XII is a serine protease derived from the liver and involved in the early stages of contact activation of the coagulation cascade with additional roles in complement activation, fibrinolysis, and the kinin system. Levels of Factor XII correlate with some features of the metabolic syndrome including triglyceride [32]; as with Factor VII there is evidence that triglyceride activates Factor XII.

**Fibrin structure in diabetes**

Activation of the coagulation cascade and the generation of thrombin with formation of cross-linked fibrin is an important component of occlusive thrombus. Increased levels of coagulation factors, particularly fibrinogen, with enhanced thrombin generation, will tend to produce a fibrin structure that is denser and more tightly packed [14]. This structure, independent of the changes in levels of lytic proteins described earlier, is more resistant to lysis and is less elastic, generating a rigid thrombotic mesh. In addition to the effects of coagulation proteins, studies in subjects with diabetes using purified fibrinogen, clearly indicate that post-translational modifications to the fibrinogen molecule, which could include glycation, oxidation or phosphorylation also affect clot structure adversely. Clots formed from purified fibrinogen of individuals with T2DM show reduced permeability, indicating more compact clots, which correlated with HbA1c levels and hence glycemic control [33]. Jörneskog reported changes in clot structure in T1DM subjects showing reduced permeability to indicate a more compact structure [34]. Furthermore, fibrin generated from purified fibrinogen in subjects before and after improvements in glycemic control demonstrates reversal of the adverse changes in fibrin structure/function and although this might superficially point to glycation as the cause, this is difficult to establish definitively [35]. Molecular studies of purified fibrinogen from subjects with T2DM indicate that the predominant abnormality associated with the change in clot structure is inhibition of fibrinolysis. Under these circumstances the equilibrium binding affinity for both tPA and plasminogen with fibrin was reduced and as a consequence, plasmin generation on the clot surface was markedly impaired. In addition, FXIII-induced cross-linking of antiplasin to fibrin was also enhanced tending to further inhibit plasmin generated on the clot surface [36]. The evidence indicates that post-translational modifications to fibrin(ogen) are promoting structural alterations to fibrin [37]. However, such changes facilitate decreased plasmin generation on the clot surface and increased antiplasin binding with an overall effect on fibrinolysis rather than clot formation [36]. Compact structures have been associated with increased cardiovascular risk and poorer cardiovascular outcome in nondiabetic populations and it is likely that a range of metabolic influences affect this phenotype.

**The platelet in diabetes**

The circulating platelet is sensitive to a wide range of metabolic changes associated with diabetes. Hyperglycemia increases platelet reactivity and improvement in glycemic control ameliorates these effects. It has been proposed that hyperglycemia may have osmotic effects on the platelet, alter protein kinase C expression and/or have indirect effects through exposure to glycated proteins. In this respect, evidence indicates that AGE proteins induce a prothrombotic state through interactions with the platelet CD36 receptor mediated by a JNK2 pathway [38]. Oxidized LDL is reported to activate platelets in insulin-resistant subjects [39] and CD36 is involved in platelet activation through interactions with dyslipidemia and oxidative stress, effects which are absent in CD36 null mice [40]. It is of note that the macrophage CD36 receptor is established as having a role in the formation of early fatty streaks through interactions with oxidized LDL leading to increased foam cell formation. This response in macrophages is accentuated in insulin-resistant states and ameliorated by thiazolidinediones, agents that are reported to possess antiplatelet effects, although it is not known whether this effect is mediated through platelet CD36. Other potential influences include effects of insulin resistance. Insulin has anti-aggregatory effects in platelets from insulin-sensitive subjects and emerging data indicate that IGF-1 may have prothrombotic effects on the platelet through interaction with the hetrodimerized insulin/IGF-1 receptor in insulin-resistant states. In a population of 208 T2DM patients with stable coronary artery disease followed up for 24 months, carriers of a particular insulin receptor substrate-1 (IRS-1) genotype exhibited both increased platelet reactivity and a significantly higher risk of major adverse cardiovascular events [41]. These findings both implicate the insulin signaling pathway in cardiovascular outcomes and provide a potential mechanism for interindividual differences between subjects with diabetes.

Overall the available data indicate that diabetes is associated with a range of metabolic abnormalities that adversely influence platelet function. Management of the platelet aspect of
Increased thrombosis risk in diabetes

**Mechanisms**

- **Impaired fibrinolytic efficiency**
  - ↑ PAI-1 levels
  - ↑ TAFI levels
  - ↑ C3 levels
  - Deranged plasmin generation and activity

- **Higher thrombotic protein levels and activity**
  - ↑ FVII levels/activity
  - ↑ vWF levels
  - ↑ FXII levels
  - Qualitative and quantitative Changes in fibrinogen

- **Enhanced platelet activation**
  - Osmotic changes
  - ↑ Protein kinase expression
  - Interaction of glycated proteins with CD36
  - Loss of platelet inhibitory effects of insulin
  - Possible role for ↑IGF-1

Figure 72.2  Pathways leading to increased thrombosis risk in diabetes. A thrombotic environment in diabetes results from enhanced platelet activation, increased coagulation factor levels and activity coupled with impaired fibrinolysis, by mechanisms that involve various proteins as shown. PAI, plasminogen activator inhibitor; TAFI, thrombin-activatable fibrinolysis inhibitor; C3, complement C3; vWF, von Willebrand factor; IGF, insulin-like growth factor.

This prothrombotic state should involve normalization of the metabolic changes seen in diabetes and the appropriate use of antiplatelet therapy as discussed later.

The mechanistic pathways responsible for increased thrombosis potential in diabetes are summarized in Figure 72.2.

**The effects of glucose and glycation on thrombosis**

The progressive nature of both T1 and T2DM tends to culminate in common metabolic changes which include insulin resistance, varying degrees of β-cell failure, and hyperglycemia, which directly affect thrombosis potential and which have been recently reviewed [42]. Attempts to dissect out the relative contributions of hyperglycemia and hyperinsulinemia to thrombotic phenotypes have been carried out in vivo using various clamp techniques and in vitro by assessing effects on proteins and cell systems involved in thrombosis. Employing hyperglycemic clamps, it has been reported that there are associated increases in thrombin–antithrombin complexes and soluble tissue factor [43]. In the same study hyperinsulinemic euglycemia was associated with increases in PAI-1. The latter findings are in agreement with some studies which report increased synthesis and secretion of PAI-1 from cells stimulated with insulin, findings which are not, however, consistent with all in vitro or in vivo studies. There are suggestions that LDL- and VLDL-induced increases in PAI-1 are augmented by glycation of apoproteins [44], and this could potentially explain some of the inconsistencies found in clinical studies if similar mechanisms applied to the effects of insulin stimulation. As described earlier, both plasmin generation and fibrin structure are modified by glycation with further reductions in lysis and the generation of a fibrin structure that is resistant to lytic degradation. Improvements in glycemic control enhance plasmin generation [26] and a combination of improved glycemic control and vitamin C reductions in oxidative stress is reported to ameliorate increased thrombin generation in T1DM [45]. Both hyperglycemia and glycation appear to have a role in monocyte and platelet activation, important components of the cellular phase of thrombosis. In nondiabetic subjects, platelet CD40L expression was increased by hyperglycemia as were the prevalence of monocyte/platelet aggregates and monocyte tissue factor expression in the presence of combined hyperglycemia/ hyperinsulinemia [46]. A potential mechanism for the direct effects of glucose on platelets comes from identification of glucose-regulated aldose reductase stimulation of platelet thromboxane which would potentially have pro-aggregatory effects on the platelet in vivo [47]. Glycation also increases tissue factor expression in both monocytes and endothelial cells and studies in the Badiomon chamber, which measures in vivo thrombosis by passing blood over a porcine artery, support the concept that changing levels of glycation alter tissue factor expression [48]. Overall, the evidence indicates that the metabolic milieu associated with diabetes, including hyperglycemia, glycation, oxidative stress, and hyperinsulinemia have the capacity to interact at all levels with the fluid and cellular phases of thrombosis to increase thrombotic risk and the clinical presentation of cardiovascular disease.
**Inflammatory, thrombotic interactions**

Multiple strands of evidence link inflammation and thrombosis as interrelated systems that are physiologically active in coordinating responses to trauma to prevent bleeding and infection. Cellular components of both systems (monocyte/macrophage and platelet/endothelium) share common pathways of activation, features of which are common to diabetes (hyperinsulinemia, glucose/glycation, dyslipidemia, oxidative stress) and there is substantial evidence indicating that inflammation/thrombosis is important in the pathogenesis of both diabetes [49] and cardiovascular disease [50]. Platelets have an important role linking inflammation and thrombosis interacting with inflammatory cells and the vascular endothelium [51]. The monocyte/macrophage forms aggregates with platelets activated at the site of vascular injury which leads to the release of a range of soluble inflammatory/thrombotic factors that contribute to hemostasis and thrombosis [52]. Increased monocyte/platelet aggregates occur in subjects with coronary artery disease [53] and are a feature of diabetes. In addition to the cellular aspects, a range of inflammatory cytokines promote release of prothrombotic elements, including PAI-1, to enhance thrombus formation and inhibition of clot lysis [54]. At the level of clot formation, proteomic analysis of fibrin clots demonstrates binding of a wide range of around 20 inflammatory proteins, including complement C3, factor H, and carboxypeptidase N catalytic chain [13]. Complement C3 binds avidly to both fibrinogen and fibrin and prolongs clot lysis, an effect exacerbated by the presence of diabetes [25]. Key interactions between the thrombotic and inflammatory pathways are highlighted in Figure 72.3.

**Circadian regulation of thrombotic mechanisms**

Regulation of metabolic and cellular functions in relation to day/night exposure (circadian variation) and the seasons (circannual variation) is a fundamental biologic feature of life on earth. The central hypothalamic and peripheral clock regulate a wide range of physiologic processes including the cardiovascular, metabolic, hemostatic, and reproductive systems in response to light and food intake to integrate function with the varying requirements imposed by a circadian life. It has been proposed that disruption of circadian responses in man secondary to chronic obesity and loss of daily and seasonal circadian cues is related to the increased risk of developing both diabetes and cardiovascular disease upon adopting Western lifestyles [55]. The central clock is comprised of a series of activators and inhibitors which set up cyclical expression in response to variation in light exposure and which generate core outputs that regulate peripheral tissue clocks. The first recognized component of the clock was termed Clock and this protein forms a heterodimer with BMal1, to activate other core clock components including BMal 2, Cry1, Cry2, Per1, and Per2. Mutations in the Clock gene have been associated with a diabetes phenotype in mice and with the metabolic syndrome [56] and sleep disorders in man. Aronson suggested that abnormal circadian rhythms in diabetes may be responsible for increased thrombotic risk [57] and since then a body of work has appeared to support this view. At around the same time, it became apparent from in vitro studies that the Clock/BMal heterodimer regulated PAI-1 gene expression through interactions with the PAI-1 gene E-box [58]. Murine studies have since confirmed the importance of Clock in fibrinolysis; Clock knockout, whether by disruption of the gene [59] or by sRNA [60] is associated with a fall in plasma PAI-1 levels and shortening of the euglobulin clot lysis time to indicate enhanced fibrinolysis. Cheng et al. [60] reported an associated increase in FVII with prolongation of the prothrombin time and activate partial thromboplastin time to suggest an additional antithrombotic effect, whilst Ohkura et al. reported no changes in prothrombin time, FVII, FX or fibrinogen. Several studies have implicated obesity in disorders of circadian rhythm and in 2006 Oishi [59] reported that Clock was involved in the regulation of PAI-1 in obesity, opening the door for a link between obesity, diabetes and cardiovascular disease through circadian rhythms and PAI-1 gene expression. Evidence to support this was also provided in animal studies implicating Clock in the upregulation of PAI-1 associated with diabetes in streptozotocin-induced murine models [61]. Studies of other core clock genes have reported that the BMal1 murine knockout is associated with increased vWF, fibrinogen, and PAI-1 [62] and that Per 2 attenuates obesity-related inhibition of fibrinolysis by reducing PAI-1 in a BMal-dependent manner [63]. Cry1 and Cry2 knockout mice had no effect on fibrinolysis or thrombotic measures [64].

![Figure 72.3 Interactions between the thrombotic and inflammatory systems.](image-url)
Although generally less well studied, the work that has been carried out on platelets shows a similar circadian rhythm to that seen for fibrinolysis. In humans, platelet count, activity and platelet markers all show significant diurnal variation with peak levels of activity in the morning [65] corresponding to the commonest time for myocardial infarction. In murine studies, Clock knockout is associated with abolition of diurnal variation in platelet activity and a reduction in platelet numbers to suggest an important physiologic role for Clock in the cellular phase of thrombosis [66]. Per2 knockout mice had almost 50% of normal platelet counts in the peripheral blood and showed decreased aggregation [67]. Studies of the effects of diabetes and/or obesity on circadian regulation of platelet function have not yet been reported.

Ameliorating thrombotic risk in diabetes

Hypoglycemic agents
The importance of insulin resistance in the early manifestations of thrombotic risk is emphasized by studies of insulin sensitizing agents which consistently demonstrate that metformin and thiazolidinedione use is associated with reduction in PAI-1 and tPA plasma levels, whilst metformin has additionally been reported to lower fibrinogen, Factor VII and XI/IIIa levels and to have direct, potentially beneficial effects on fibrin structure [14]. The most compelling evidence for the longer term benefits of insulin sensitization on thrombotic risk comes from the BARI-2D trial [68]. In this study, over 2300 T2DM patients with stable coronary artery disease were randomized to either an insulin-sensitizing (metformin or glitazone) or an insulin-providing regimen and followed up for 5 years. Striking reductions in PAI-1, tPA, C-reactive protein and fibrinogen were seen at all points in patients treated with insulin sensitizers during the study whilst changes in D-dimer and fibrinopeptide A (markers of clot formation and degradation) were similar in the two groups [69]. Baseline values of fibrinogen, D-dimer and C-reactive protein but not PAI-1 or tPA were strongly related to predesignated vascular outcomes. Insulin providers, including the sulfonylureas, have beneficial effects on glucose-mediated alterations in thrombosis; there is as yet very little information on the effects of drugs affecting the incretin system.

Aspirin
In addition to the antiplatelet effects, aspirin acetylates fibrinogen, altering fibrin characteristics [70–72] and may influence clot lysis indirectly through a nitric oxide-dependent mechanism [73,74]. These platelet-independent fibrinolytic properties of aspirin may explain the enhanced fibrinolytic effects of streptokinase when used with aspirin [75]. Aspirin is regularly used in the setting of acute coronary syndrome (ACS), the benefit of which has been repeatedly demonstrated in individuals with or without diabetes [76,77]. Longer term, aspirin is used for secondary cardiovascular protection in diabetes [78,79], a practice supported by two large meta-analyses [80,81]. The use of aspirin for primary cardiovascular protection in diabetes is more controversial [81,82], with two recent primary prevention studies failing to show an impact of aspirin in diabetes [83,84]. A small study of 651 diabetes subjects over 11.6 years follow-up, reported a reduction in CV events in aspirin-treated subjects after adjustment for significant CV variables, indicating that aspirin may be beneficial in some patients with diabetes [85]. In contrast, an increase in cardiovascular events was reported in aspirin-treated Chinese and Swedish diabetes subjects with no history of ischemic heart disease [86,87]. The reasons for these findings are not clear but may be related to study design (both observational and nonrandomized) and possible antiplatelet treatment of higher risk subjects. Overall, data indicate that the efficacy of aspirin in primary prevention in diabetes is compromised and should be used only in higher risk individuals. Future studies are warranted to fully categorize diabetes subjects who benefit from aspirin therapy for primary cardiovascular prevention.

Clopidogrel
Clopidogrel is a thienopyridine that irreversibly inhibits the platelet P2Y12 receptor. Clopidogrel is a prodrug converted to the active metabolite by the P450 system in the liver and onset of action may be delayed by CYP genetic polymorphisms. Clopidogrel is used in combination with aspirin in subjects with acute coronary syndrome and as monotherapy in those intolerant to aspirin or in patients with symptomatic cerebrovascular disease despite aspirin therapy [88,89].

The combination of aspirin and clopidogrel following ACS has been regularly used in clinical practice supported by a number of studies, with benefits demonstrated in both diabetic and non-diabetic subjects [90–93]. However, newer agents have recently shown a superior efficacy with centers moving gradually away from clopidogrel therapy (detailed later).

Prasugrel
Similarly to clopidogrel, this agent irreversibly blocks the P2Y12 receptor but has a theoretical advantage through quicker metabolism and faster onset of action. Although prasugrel has shown a better efficacy than clopidogrel at reducing ischemic events following ACS, the increase in bleeding risk wiped out the benefit [94]. However, subgroup analysis of the diabetes group showed a benefit without an increase in bleeding risk, an effect that was particularly pronounced in insulin users. This suggests that diabetes subjects have a different response to antiplatelet agents and may require different antithrombotic therapy compared with the nondiabetic population.

Ticagrelol
Ticagrelol blocks the platelet P2Y12 receptor; however, it differs from the thienopyridines described earlier by (i) being an active compound, (ii) reversibility of action, and (iii) shorter half-life necessitating twice daily administration. The PLATO
Hemostatic abnormalities in diabetes mellitus

A trial showed superior efficacy of ticagrelol compared with clopidogrel when used in combination with aspirin in 18,642 ACS patients treated medically or following PCI [95]. This agent was equally effective in individuals with diabetes [96], and therefore is becoming the antiplatelet of choice to use in combination with aspirin following ACS in some centers.

**Dipyridamol and cilostazol**

These agents modulate the phosphodiesterase pathway to reduce platelet activation. Given the absence of an indication for dipyridamol in coronary artery disease and the questionable efficacy in cerebrovascular disease, this agent is not used frequently in clinical practice. In contrast, the use of cilostazol appears to be gaining momentum in diabetes, although large-scale studies are lacking [97,98].

**Inhibitors of platelet–fibrinogen interaction**

A meta-analysis of GPIIb/IIIa inhibitors, including tirofiban, lamifiban, etifibatide, and abciximab, suggested superior efficacy of these agents in subjects with diabetes undergoing PCI [99]. However, further studies have given conflicting results in both diabetic and nondiabetic subjects, making routine use of these agents difficult to justify [100,101].

**Coagulation inhibitors**

Heparin molecules are indirect inhibitors of FX and prothrombin, through modulation of antithrombin III activity. Fractionated heparin is used in ACS, including diabetes. Enoxaparin is the main low molecular weight heparin used because of the predictive anticoagulative effect, ease of injections, and lower risk of thrombocytopenia.

Bivalirudin is a direct thrombin inhibitor recommended for diabetes subjects with ACS in whom coronary intervention is planned, particularly those with a high bleeding risk [102,103]. More work is needed to clarify the characteristics of individuals who would gain most benefit from this therapy. Table 72.1 summarizes the main agents that can influence long-term thrombosis risk in diabetes.

---

**Summary and conclusions**

Diabetes mellitus is a chronic condition characterized by the presence of hyperglycemia and an increased risk of cardiovascular disease. Although T1 and T2DM have a distinct pathologic basis, the progressive nature of these disorders tends to lead to convergent characteristics over many years following diagnosis. For example, individuals with T1DM can become insulin-resistant, displaying features of T2DM, whereas individuals with T2DM may progress to β-cell failure, requiring insulin therapy. Inflammatory thrombotic interactions are fundamental to the generation of vulnerable vascular lesions and to the formation of an occlusive thrombus when plaques become unstable and rupture. These interactions are sensitive to a range of metabolic abnormalities associated with both insulin resistance and hyperglycemia/glycation and the progressive nature of diabetes is tracked by progressive changes in risk for vascular damage. Thrombotic processes include the fluid and cellular phases of coagulation, fibrinolysis, and platelet function and all aspects are affected in a potentially deleterious way by diabetes and developing complications. It is important to recognize that occlusive vascular damage occurs over many years and involves a wide range of pathophysiologic processes of which thrombosis is only a part. Just as complex interactions generate a vascular risk profile, so therapeutic interventions that deal with multiple risk factors tend to ameliorate risk. In this respect management of insulin resistance through weight loss, exercise, insulin-sensitizing agents, glycemia/glycation, and lipid lowering will all improve multiple risk factors in addition to their effects on thrombosis. In the presence of acute coronary syndromes the use of antiplatelet agents, direct thrombin inhibitors, and other approaches affecting thrombotic

---

**Table 72.1** The role of different long-term treatment agents in ameliorating the thrombosis risk in diabetes. It should be noted that agents that reduce thrombosis risk can increase the risk of bleeding, whereas hypoglycemic agents can both reduce and increase thrombosis risk depending on blood glucose levels

<table>
<thead>
<tr>
<th>Agent</th>
<th>Benefits</th>
<th>Risks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypoglycemics</td>
<td>Reduction of protein glycation</td>
<td>Hypoglycemia (which can be prothrombotic)</td>
</tr>
<tr>
<td>Aspirin</td>
<td>Reduction in platelet aggregation</td>
<td>Bleeding complications</td>
</tr>
<tr>
<td>Clopidogrel</td>
<td>Enhanced fibrinolysis</td>
<td>Reduced efficacy in diabetes (unclear mechanisms)</td>
</tr>
<tr>
<td>Prasugrel</td>
<td>Reduction in platelet aggregation</td>
<td>Bleeding complications</td>
</tr>
<tr>
<td>Ticagrelol</td>
<td>May be more effective in diabetes</td>
<td>Reduced efficacy in some individuals</td>
</tr>
<tr>
<td>Cilostazol</td>
<td>Good safety profile</td>
<td>Cost Dual therapy only</td>
</tr>
<tr>
<td>Ticagrelol</td>
<td>Reduction in platelet aggregation</td>
<td>Bleeding complications</td>
</tr>
<tr>
<td>Cilostazol</td>
<td>Reduction in platelet aggregation</td>
<td>Cost Dual therapy only</td>
</tr>
</tbody>
</table>

**Ameliorating the thrombotic environment in diabetes using oral therapy**
processes has radically altered outcomes in both diabetic and nondiabetic subjects.

Perhaps the future in this rapidly changing field lies in a clear understanding of the relationship between obesity-induced changes in circadian rhythm and alterations in thrombotic risk in particular and vascular risk in general. The potential implications for how we live our lives and for ameliorating the effects of modern lifestyles on the current epidemic of diabetes and its complications are currently difficult to estimate. This is likely, however, to continue to be a major area of interest in the coming years.

References

concentration dependent manner: a mass spectrometric and isotope labeling study. Biochemical and Biophysical Research Communications 2012;421:335–342.


95 James S, Akerblom A, Cannon CP, et al.: Comparison of ticagrelor, the first reversible oral P2Y12 receptor antagonist, with clopidogrel in patients with acute coronary syndromes: rationale, design, and baseline characteristics of the PLATelet inhibition


CHAPTER 73

Clinical features and treatment of coronary heart disease in diabetes

Eberhard Standl¹, Michael Erbach², Markolf Hanefeld³, Oliver Schnell¹, Bernd Stratmann⁴ and Diethelm Tschoepe⁴

¹Munich Diabetes Research Group e.V. at Helmholtz Centre, Munich-Neuherberg, Bavaria, Germany
²Sciarc Institute, Baierbrunn, Germany
³Gesellschaft für Wissens- und Technologietransfer mbH at Technical University Dresden, Dresden, Saxony, Germany
⁴Heart and Diabetes Center NRW, University Clinic Ruhr University, Bochum, Germany

Key points

- Cardiovascular disease develops frequently in patients with diabetes and is a cause of death in some 70% of these patients.
- Conversely, some 70% of patients with chronic or acute coronary heart disease have diabetes—known or undiagnosed—or prediabetes (impaired glucose tolerance); hence, all patients with established coronary heart disease without known diabetes should be screened for hidden diabetes.
- Cardiac autonomic neuropathy, (functional) microangiopathy, hypoglycemia, cardiomyopathy, and heart failure are specific features in patients with diabetes and impact on the symptomatology, complexity, and prognosis of heart disease in these patients.
- Serious cardiac arrhythmias may occur in the context of high glucose variability or excessive hyper- and hypoglycemia.
- Appropriate intensive multifactorial medical therapy to reduce total cardiovascular risk and consequent guideline-based revascularization therapy are the cornerstones in the management of patients with diabetes and coronary heart disease.

Introduction

Diabetes mellitus and coronary heart disease (CHD) often appear as two sides of the same coin. On the one side, diabetes remains as an independent cardiovascular (CV) risk predictor that still today doubles the risk of CV morbidity and mortality. Some 70% of all patients with diabetes will ultimately die from CV disease (CVD), mostly from cardiac complications, despite all the progress made in interventional and conservative cardiology in recent years [1,2]. On the other side, some 70–75% of all patients with established CHD present with diabetes or significant perturbations of glucose homeostasis.

With a growing tendency, more than 30% of contemporary cardiology patients exhibit coexisting known diabetes. In addition, as many as 10–15% of cardiology patients have previously undiagnosed diabetes and another quarter has impaired glucose tolerance (IGT) or prediabetes, as the Euro Heart Survey [3] and other epidemiologic evaluations have shown. Thus, there is a huge overlap between the two fields of diabetology and cardiology warranting joint action, and in 2007 the first-ever cardio-diabetologic guidelines on diabetes, prediabetes, and cardiovascular diseases were launched by the European Society of Cardiology and the European Association for the Study of Diabetes [4]. Similar joint guidelines have since followed on other continents of the world. This chapter attempts to give a comprehensive summary of the joint cardio-diabetologic approach, describing the clinical features and the treatment of CHD from the perspective of the diabetologist.

Key practical features of the relationship between CHD and diabetes

CHD at onset of diabetes

CHD may be present at diagnosis of T2DM, as found, for example, in the Munich General Practitioner Project. Some 40% of patients showed CHD-related ECG changes. The reasons for this observation seem to be complex and include, for example, a potential long lag time between the clinical onset and diagnosis of T2DM or clustering of other CV risk factors such as hypertension or dyslipidemia with T2DM—also seen in the Munich GP Project [5]. This latter phenomenon is often referred to as the metabolic syndrome and originates in the dysglycemic diabetic pre-state below the threshold of T2DM. Epidemiologic evaluations have shown that the risk for CV events is several
fold increased up to 15 years prior to the diagnosis of diabetes. Furthermore, when oral glucose tolerance tests have been applied, the state of glucose intolerance and—to a lesser degree—the state of impaired fasting glucose were found to be associated and highly predictive not only for CV events and mortality, but also all-cause mortality [6,7]. The association with fasting glucose may even show a J-type curve in that very low glucose concentrations at the far low end are also associated with increased all-cause and CV mortality [8]. Very low fasting glucose, however, may not necessarily be a causative factor for CV mortality, but may rather be a surrogate of high CV risk as associated with confounding comorbidities and diseases.

**Duration of diabetes and CHD**

Notwithstanding the fact that CHD may be present at diagnosis of T2DM, there is also an excessive rise of CV events and mortality with duration of diabetes, both type 1 and 2, in comparison to age-matched nondiabetic cohorts [9]. Looking into gender differences, the relative risk of women with diabetes for CVD as compared to women without diabetes is significantly higher than in diabetic versus nondiabetic men [9]. In longitudinal studies starting around 1990, after some 20 years of diabetes, about 40–50% of patients of both types and sexes have died from CV causes, in contrast to only about 7% of women and 16% of men without diabetes [1]. On the other hand, since the 1970s a secular trend for a reduction of CV and CHD mortality by about 70% has also occurred in people with diabetes, probably due to the progress in CV risk factor management and interventional cardiology [2]. Diabetes today is no longer an equivalent of CHD (i.e., a 10-year CV mortality risk of more than 20%, as stated in the 1990s based on cohorts studied in the 1980s). In fact, the CV mortality risk depends largely on coexisting CV risk factors and seems to be down well below 10% over 10 years on average in contemporary cohorts [10]. However, diabetes still carries a twofold elevated CV mortality risk compared to nondiabetic subjects.

**Acute coronary syndromes and diabetes**

The enormous interaction of diabetes and CHD is exquisitely visible at the interface of acute coronary syndromes (ACS) and hyperglycemia/diabetes. Random blood glucose concentration on admission with ACS is highly predictive not only for in-hospital mortality, but also for the prognosis longer term: the higher the level, the poorer the outcome. The striking prevalence (a total of 70–75%) of coexisting known diabetes, previously undiagnosed overt diabetes, and prediabetes in patients with ACS and CHD has already been mentioned. Both the Euro Heart Survey on Diabetes and the Heart, and the German SWEETHEART Registry have looked into this in some detail [3,11]. The poorest prognosis in terms of mortality and MACCE (major adverse cardio- and cerebrovascular events, i.e., deaths and nonfatal MIs and strokes) over the subsequent years is seen in those ACS patients with known diabetes, followed by those with new diabetes; with a trend for a poorer outcome in those with prediabetes emerging at 3-year follow-up compared to nondiabetic controls. A similar “dose–effect” of known, new, and prediabetes has also been observed in the context of PCI (percutaneous coronary intervention). Treating newly diagnosed diabetes in ACS patients appeared to be beneficial in observational studies longer term, for example in the Euro Heart Survey at 1-year follow-up [12], whereas the benefit of an acute attempt to normalize glycemia on admission with ACS has remained controversial after studies like DIGAMI 1 and 2, CREATE-ECLA, and others. It seems reasonable to target a blood glucose concentration of about 150 mg dL\(^{-1}\) in an acute situation, if necessary with insulin, but to avoid any degree of hypoglycemia. The latter seems to signal a bad prognosis, also in those without diabetes and insulin. It is hoped that once blood glucose-lowering drug therapies without risk of hypoglycemia, for example incretin-based therapies, have been approved for use in ACS, more robust recommendations for dealing with hyperglycemia in ACS will become available.

**Prognosis of CHD and diabetes**

The prognosis of CHD in diabetes largely depends on coexisting CV risk factors and additional confounding conditions such as more severe impaired kidney function, cardio-neuropathy, complicating heart failure, severe lung or liver dysfunction, inflammatory disease, and frailty of any origin [10]. As already touched upon, however, the extent of hyperglycemia—best measured as HbA1c—was also found to be highly predictive for macrovascular, especially cardiac fatal and nonfatal events in numerous long-term observational studies. A 1% increase of HbA1c seems to be associated with an 18% increase of incident CV complications. In the Munich GP Project looking at mortality over 10 years, the association with initial HbA1c concentration was also confirmed in the subgroup that were already diagnosed at baseline as having significant macrovascular disease [13]. Based on epidemiologic judgment, including the observational data from recent large-scale trials such as ACCORD [14] and ADVANCE [15], an attained mean HbA1c level of 7% and less is seen as helpful to minimize macrovascular events, provided those levels can be achieved safely and without hypoglycemic episodes.

Hypoglycemia has emerged as an important factor, in the context of cardiovascular mortality, which complicates blood glucose-lowering therapy (see also later in this chapter). In three of the recent randomized and controlled intervention trials (RCTs) evaluating more intensive blood glucose-lowering treatment, hypoglycemic events were a strong predictor of CV mortality. This does not necessarily mean causality, but may indicate a high CV risk situation due to other coexisting disease. At any rate, hypoglycemia was felt to play a major role in why more intensive blood glucose-lowering therapy yielded only mixed CV outcome results in RCTs, for example in ACCORD [14], ADVANCE [15], and VADT [16]. It seems noteworthy in this context that severe hypoglycemia also occurred in the less intensively treated group and in association with increased CV...
mortality, whereas the intensive treatment group, especially those patients with the highest updated HbA1c levels, showed the highest mortality risk. In the UKPDS, the beneficial effect of more intensive therapy in reducing myocardial infarction was of borderline significance after 10 years of this RCT, and only became significant after 10 years’ further observation as a “legacy effect” [17]. Similarly, DCCT (a trial in T1DM) produced a legacy effect in terms of preventing CV events during the open observation period after the end of the RCT [18]. ORIGIN, which besides overt T2DM included also newly diagnosed diabetes and prediabetic subjects at high CV risk or with coexisting CVD (some 60% of participants), generated only an HbA1c difference of 0.3% over time, and did not demonstrate a more favorable CV outcome in patients with more intensive glycemic control [19]. Conversely, several meta-analyses based on UKPDS, ACCORD, ADVANCE, and VADT showed a significant reduction of nonfatal myocardial infarction and of all cardiac events of some 15–17% associated with a mean HbA1c reduction of 0.9% over a mean follow-up time of 5 years [20]. Looking at the subgroup with pre-existing macrovascular disease, however, no benefit of more intensive blood glucose-lowering treatment could be substantiated in these meta-analyses.

So, in aggregate, especially in patients with coexisting CVD, an individualized, patient-centered blood glucose-lowering approach is warranted. Care should be taken, not to do harm, and when addressing HbA1c lowering, always start slow, do not be overaggressive, go slow. An HbA1c of 7% or even lower may be fine, as long as there are no discernible side effects. In case of a baseline HbA1c above 8%, a reduction below 8% seems to be a reasonable first step (as a safeguard to prevent more severe microvascular disease). Further HbA1c reduction should be advised only if it can be done safely, that is, without inducing hypoglycemia.

To achieve HbA1c goals, recommendations for glycemic targets are also needed. The ORIGIN study has shown that targeting fasting plasma glucose concentrations of 95 mg dL\(^{-1}\) is safe and associated with only a low rate of severe hypoglycemia in about 1/100 patient-years. As to postprandial glucose levels and their predictive impact of increasing CV complications in long-term studies, a target glucose concentration of 180 mg dL\(^{-1}\) two hours after a meal seems to be a reasonable goal and should ensure an HbA1c result of about 7%. Key characteristics of the five studies evaluating intensification of blood glucose-lowering therapy are summarized in Table 73.1.

**Screening for diabetes and prediabetes in patients with established CHD**

Three good reasons seem to underpin the recommendation to screen for diabetes and prediabetes in patients with established CHD: (1) the high prevalence of undiagnosed diabetes and prediabetes in patients with both chronic and acute CHD; (2) the associated adverse impact of these conditions on the further prognosis; (3) the potential of improving the prognosis by appropriate therapeutic measures. As surveys such as the Euro Heart Survey and the SWEETHEART Registry of the Silent Diabetes Study have shown, performing an oral glucose tolerance test has remained the assessment of choice to diagnose these hidden metabolic disorders, notwithstanding the fact that well-standardized HbA1c measurements have been approved for the diagnosis of diabetes [3,21]. Both measurements of fasting plasma glucose and HbA1c, however, have been found to miss the diagnosis of two-thirds of all cases with overt diabetes as detected by an oral glucose tolerance test (OGTT) [3,21]. Furthermore, impaired glucose tolerance can only be detected by an OGTT per definition.

In practical terms, it seems useful to arrange for an examination of the glycemic status in all patients with established CHD, unless diabetes is already known. This can be done in the context of coronary angiography or other imaging procedures of the heart for patients with chronic CHD, and for patients with ACS at the end of their stay in the Coronary Care Unit and after stabilization of the patient’s circulation. To minimize the

Table 73.1 | Synopsis of long-term, outcome-oriented RCTs looking into intensive blood glucose-lowering therapy [14–17,19,20]

<table>
<thead>
<tr>
<th>Study</th>
<th>n</th>
<th>Age (yrs)</th>
<th>Diabetes (yrs)</th>
<th>Pre-diabetes (%)</th>
<th>Macrovasc. compl. (%)</th>
<th>Baseline HbA1c (%)</th>
<th>Intensive Rx target</th>
<th>Intervention</th>
<th>Median follow-up (yrs)</th>
<th>Mean HbA1c during follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>UKPDS</td>
<td>4620</td>
<td>53</td>
<td>&lt;1</td>
<td>0</td>
<td>NS</td>
<td>7.1</td>
<td>FPG ≤ 6 mmol L(^{-1})</td>
<td>SU or Insulin or metformin, plus multiple drugs as necessary</td>
<td>10.1</td>
<td>7.9%</td>
</tr>
<tr>
<td>ACCORD</td>
<td>10,251</td>
<td>62</td>
<td>10</td>
<td>0</td>
<td>35</td>
<td>8.1</td>
<td>A1c &lt; 6%</td>
<td>Multiple drugs</td>
<td>3.5</td>
<td>7.5%</td>
</tr>
<tr>
<td>VADT</td>
<td>1791</td>
<td>11.5</td>
<td>0</td>
<td>40</td>
<td>9.4</td>
<td>A1c &lt; 6%</td>
<td>Multiple drugs</td>
<td>5.6</td>
<td>8.4%</td>
<td></td>
</tr>
<tr>
<td>ADVANCE</td>
<td>11,140</td>
<td>8</td>
<td>0</td>
<td>32</td>
<td>7.5</td>
<td>A1c ≤ 6.5%</td>
<td>Gliciazide ± others</td>
<td>5.0</td>
<td>7.3%</td>
<td></td>
</tr>
<tr>
<td>ORIGIN</td>
<td>12,612</td>
<td>5</td>
<td>19</td>
<td>66</td>
<td>6.5</td>
<td>FBG ≤ 5.3 mmol L(^{-1})</td>
<td>Gargine ± others</td>
<td>6.2</td>
<td>6.5%</td>
<td></td>
</tr>
</tbody>
</table>

*Excluded individual with current angina or heart failure, and those with more than 1 major vascular event in the past or myocardial infarction in the previous year.*
need to perform an OGTT, it has been suggested to do a combined measurement of fasting plasma glucose and HbA1c first, subsequently followed by an OGTT in cases where readings under the cut-off for overt diabetes were obtained in the first instance (see also Table 73.2).

### Specific clinical features of CHD in diabetes

#### Morphology of CHD
In patients with diabetes, accelerated coronary artery disease is common. In adult diabetic patients, a three- to fourfold prevalence of CHD as compared to the general population has been observed. In patients with diabetes, a diffuse and rapidly progressive form of atherosclerosis is common [22]. Furthermore, a higher incidence of two- and three-vessel disease has been reported [23]. Based on imaging methods such as ultrafast computed tomography or intravascular ultrasound, a greater extent of calcification and a decrease in adaptive remodeling of atheroma have been detected. In addition, an increase in macrophage infiltration, vasoconstriction, adhesion molecule activation, inflammation, coagulation, and platelet activation has been reported. In patients with diabetes, a more pronounced plaque burden and positive remodeling have been found [24]. A study based on multislice computed tomography, found a higher prevalence of obstructive atherosclerosis in patients with T2DM as compared to patients with T1DM. In addition, noncalcified plaques were more common in patients with T2 than with T1DM [25].

#### Symptomatology
Angina pectoris, dyspnea, nausea, fatigue, and light-headedness are reported as typical symptoms of cardiac disease. Due to diabetic autonomic dysfunction, symptoms may not be characteristic or may be either partially or entirely absent. Shortness of breath, sweating, or profound fatigue are often recorded complaints [4].

In the recently published "Silent Diabetes Study," a significant correlation between the prevalence of abnormal glucose regulation (impaired glucose tolerance, diabetes) as detected by OGTT and the extent of CAD has been detected [21]. The study also ties in with a prevalence of undetected diabetes of more than 20% in patients with acute myocardial infarction, which has also been found by different studies. It has been demonstrated that glucose metabolism is altered in a large proportion of patients with acute myocardial infarction and no previous diagnosis of diabetes mellitus [21].

### Cardiac autonomic neuropathy
Cardiac autonomic neuropathy (CAN) is one of the most common complications associated with diabetes. It is caused by injury to the autonomic nerve fibers that innervate the heart and blood vessels. An increased risk of mortality and major cardiovascular events has been reported to be associated with CAN [4]. Different pathogenic mechanisms are considered [26].

- Activation of the polyol pathway, which is followed by an accumulation of sorbitol and fructose, a depletion of myoinositol, and a reduction of Na+/K+-ATPase. In this way, intracellular osmotic stress is increased.
- Accumulation of non-esterified free fatty acids and advanced glycation end-products (AGEs) as a result of hyperglycemia.
- Oxygen species, produced by the mitochondrial respiratory chain, are sufficient to oxidize lipids.
- Activation of mitogen-activated protein (MAP) kinases, which contribute to a range of fundamental cellular processes such as growth, proliferation, and differentiation.
- Reduced circulation, followed by endoneural and ganglionar hypoxia, which is accompanied by myocardial hypoxia.
- Lack of neuronal growth factors and impairment of axonal re-transport.
- Autoimmune reactions against sympathetic nervous tissue (only T1DM) are linked with cardiac autonomic nervous function.

The functional and structural performance of the myocardium is strongly dependent on the balance between sympathetic and parasympathetic neural function. Exercise intolerance, orthostatic intolerance, and syncopies have been reported as typical symptoms associated with CAN. Since these symptoms usually occur late in the disease process, they are only weakly related to symptom scores in mild CAN.

Due to good sensitivity, specificity, and reproducibility, cardiovascular reflex tests are considered as the gold standard in clinical autonomic testing [27]. In patients with proliferative retinopathy a Valsalva maneuver may not be performed due to the risk of vitreous hemorrhage [27]. Heart rate variability (HRV) testing with deep breathing is the most widely used
assessment of cardiac autonomic function and has about 80% specificity [27].

ECG-based functional tests, however, have the disadvantage, that early stages of cardiac autonomic dysfunction often remain undetected. The heterogeneity of cardiac dysfunction, predominance of posterior wall region, and the extent of cardiac sympathetic dysfunction have been widely unknown up to the introduction of quantitative scintigraphic assessment of sympathetic innervation via techniques such as SPECT and PET [28].

Scintigraphically assessed cardiac sympathetic dysinnervation is frequently observed in type 2 diabetic patients. The pattern of cardiac sympathetic dysinnervation is heterogeneous, whereas the posterior myocardium is predominantly affected. Type 2 diabetic patients with ECG-based cardiac autonomic neuropathy demonstrate a more pronounced dysinnervation of the posterior myocardial region than type 2 diabetic patients without. In the patients, the posterior myocardial I-123-MIBG uptake is associated with heart rate variation at rest and during deep breathing and heart rate response to standing [29]. Furthermore, global, anterior and apical I-123-MIBG uptake is related to heart rate variation at rest and during deep breathing.

The high degree of reduced I-123-MIBG uptake in type 2 diabetic patients without ECG-based cardiac autonomic neuropathy emphasizes the view that cardiac sympathetic dysinnervation is observed before ECG-based cardiac autonomic neuropathy is diagnosed. Potential differences in uptake of I-123-MIBG between T1 and T2DM are not yet fully understood and may also be related to differences in pathophysiology between the types of diabetes.

In general, CAN is suggested to be composed of a reversible metabolic component and an irreversible structural component [30]. With increase in duration of diabetes, the reversible component is considered to decrease while the irreversible component augments, respectively [30].

Studies that did not scintigraphically exclude myocardial perfusion defects reported a substantial reduction of cardiac I-123-MIBG uptake in type 2 diabetic patients [31]. A comparative investigation of I-123-MIBG uptake with power spectral analysis of heart rate variability in T2DM demonstrated a relationship between I-123-MIBG uptake and parameters of power spectral analysis. In type 2 diabetic patients with silent myocardial ischemia, diffuse abnormalities of I-123-MIBG uptake have been demonstrated.

The predominant affection of the posterior myocardium of type 2 diabetic patients points at a particular vulnerability of the posterior cardiac sympathetic nervous system. A comparable heterogeneous pattern of cardiac sympathetic dysinnervation has been observed in arrhythmogenic right ventricular cardiomyopathy and in patients with ventricular arrhythmias in the absence of coronary artery disease. Therefore, the regional imbalance of cardiac sympathetic innervation in T2DM may enhance the potential for arrhythmogenicity.

In type 2 diabetic patients, heart rate variation is significantly associated with global and regional distributions of myocardial uptake of I-123-MIBG. Investigation of 24-h heart rate variability demonstrated a relationship with I-123-MIBG uptake in type 2 diabetic patients without evidence for coronary artery disease. Cardiac reflex tests assessing heart rate variation have been suggested to mainly reflect cardiac parasympathetic function.

Tests reflecting mainly cardiac sympathetic function such as Valsalva maneuver and systolic blood pressure response neither correlated with global nor with regional I-123-MIBG uptake. Applying power spectral analysis of heart rate variability, evidence for a relationship between low-frequency components, which is believed to mainly reflect sympathetic function, and I-123-MIBG uptake has been reported in T2DM. Comparisons with scintigraphically assessed sympathetic function support the view that the tests on heart variation do not absolutely differentiate between sympathetic and parasympathetic function.

### Cardiomyopathy

Diabetic cardiomyopathy may develop independently of elevated blood pressure or coronary artery disease. This complex diabetes-associated complication is characterized by a range of molecular, structural, and functional changes (Table 73.3 [32]).

The healthy heart is able to metabolize various substrates such as free fatty acids (FA), glucose, lactate, pyruvate, ketone bodies, and amino acids. Under normal resting conditions, cardiac metabolism is mainly oxidative, and 60–90% of the ATP arise from ADP oxidative phosphorylation and fatty acid oxidation. Under stress conditions, such as ischemia or hypertrophy, oxidation of FA is decreased and exogenous glucose is mainly utilized [33,34]. The healthy heart is able to switch rapidly between different energy sources to adjust to different requirements.

As an aerobic organ, the heart is relying to a large extent on the aerobic oxidation of substrates. The myocardial oxygen consumption is closely linked to the main determinants of systolic function, namely, heart rate, contractile state, and wall stress. In pathologic states such as heart failure or cardiomyopathy, mechanical efficiency (ratio of produced useful energy to consumed oxygen) is reduced.

Due to profusion of triglycerides and non-esterified FA, the diabetic heart essentially relies on FA oxidation, despite the presence of hyperglycemia and even during myocardial ischemia. Regardless of enhanced FA utilization, FA uptake is assumed to overcharge the oxidation capacity, resulting in myocardial lipid accumulation that may promote lipotoxicity. The latter causes impaired beta-oxidation, thus generating even more free FA. Hence, an intrasarcoplasmic excess of neutral lipids and their intermediates is originated, and directed to signaling pathways whose hyperactivity then affects ATP production, insulin sensitivity, cell contractility, and apoptosis, all contributing to diabetic cardiomyopathy.

The levels of intracellular FA derivatives like fatty acyl-coenzyme A (CoA), diacylglycerol and ceramide increase with augmenting intracellular fatty acid amounts. Fatty acyl-CoA species are able to inhibit glycolysis and are presumed to regulate
Table 73.3  Natural course of diabetic cardiomyopathy [32]

<table>
<thead>
<tr>
<th>Phase</th>
<th>Molecular and cellular events</th>
<th>Structural and morphological alterations</th>
<th>Myocardial capacity</th>
</tr>
</thead>
</table>
| 1     | • Metabolic: hyperglycemia, increased circulating FFA, insulin resistance  
• Altered Ca2+ homeostasis  
• Endothelial dysfunction | • Insignificant changes: normal LV dimensions, wall thickness, and mass. | • Impaired diastolic compliance with normal systolic function, or no obvious functional changes. |
| 2     | • Cardiomyocyte injury, apoptosis, necrosis  
• Activation of cardiac fibroblasts (myocardial fibrosis) | • Minor changes: slightly increased heart mass, wall thickness or size.  
• Cardiomyocyte hypertrophy.  
• Insignificant myocardial vascular changes. | • Significant changes in diastolic and systolic function. |
| 3     | • Hypertension  
• CAD  
• Microangiopathy  
• CAN | • Significant changes: increased heart size, wall thickness and mass.  
• Myocardial microvascular disease. | • Abnormal diastolic and systolic function. |

FFA, free fatty acids; CAD, coronary artery disease; CAN, cardiac autonomic neuropathy; LV, left ventricular.

glucose uptake. Diacylglycerol activates distinct isoforms of protein kinase C, thereby eliciting insulin resistance.

Activation of these protein kinase C isoforms is assumed to be linked to glucose spikes and glucose excess with early changes of sympathetic incompetence, impaired NO release and endothelial dysfunction and activation, ultimately leading to severe damage, such as heart failure, plaque instability, arterial thrombus formation, and death.

Beyond increased FA levels, the mechanisms contributing to substrate switching in diabetic hearts include decreased insulin signaling, and activation of transcriptional pathways such as the peroxisome proliferator-activated receptor-α (PPARα)/PGC-1 signaling network that regulates myocardial substrate use. Accordingly, activation of PPARα stimulates the formation of pyruvate dehydrogenase (PDH) kinase 4, which attenuates glucose oxidation by decreasing PDH activity.

Moreover, PPARα stimulates the activation of genes such as CD36, regulating cellular FA uptake, and malonyl CoA decarboxylase, degrading malonyl CoA, thereby depressing carnitine palmitoyl transferase-1 and stimulating mitochondrial FA uptake. Further transcriptional targets of PPARα are medium- and long-chain acyl CoA dehydrogenase and hydroxyl acyl CoA dehydrogenase, both involved in beta-oxidation [35].

PPARα is highly expressed in tissues deriving most energy from FA oxidation, including liver, heart, kidney, and skeletal muscle. PPARα target genes are involved in lipid metabolism, for instance the genes of heart-type FA-binding protein (hFABP), lipoprotein lipase (LPL), CD36, carnitine palmitoyltransferase-1 (CPT-1), and uncoupling protein-3 (UCP-3). Cardiac PPARα excess enhances FA uptake and oxidation.

Increased FA metabolism not only leads to accumulation of FA and triglycerides, but also stimulates oxygen consumption and formation of reactive oxygen species (ROS). Excessive FA uptake through overexpression of LPL or FA transporters, or by stimulating PPARα expression or long-chain CoA synthase results in a cardiac phenotype equal to diabetic cardiomyopathy.

Beyond FA and lipotoxic aspects, the contribution of dysregulated carbohydrate metabolism and glucotoxicity to diabetic cardiomyopathy has to be considered. A significant reduction in myocardial glucose utilization has been demonstrated in isolated diabetic cardiomyocytes and in diabetic patients.

The slow rate of glucose transport across the sarcolemmal membrane into the myocardium as a limiting factor of cardiac glucose utilization is suggested to be caused by the cellular depletion of glucose transporters 1 and 4 (GLUT1 and GLUT4). Reduced myocardial GLUT4 content and a defect in GLUT4 translocation are the earliest changes being observed in short-term studies of mice under a high-fat diet, causing reduced rates of glycolysis and glucose oxidation.

According to the Randle hypothesis, glucose accumulates in the cells and is deflected to alternate metabolic pathways, especially the pentose phosphate pathway and the hexosamine biosynthetic pathway. Among others, increased intracellular levels of UDP-N-acetylglucosamine are suspected to alter excitation/contraction coupling by activating the O-linked glycosylation of target proteins.

Oxidative stress plays a key role in the pathophysiology of diabetic cardiomyopathy. According to recent studies, the maladaptation of the heart is based on metabolic dysfunction associated with diabetes. Key factors contributing to oxidative stress in diabetic patients in particular are [34–37]

- excess formation of ROS induced by hyperglycemia, AGE, and elevated free FA;
- reduction of antioxidant defense;
- increase in mitochondrial ROS generation;
- ROS-mediated activation of factors that are involved in the pathogenesis of diabetic cardiomyopathy: inflammation, endothelial dysfunction, cell death, cardiovascular remodeling;
• activation of transcription factors, polyol and hexosamine pathways, tyrosine kinase pathways.

Heart failure
Type 2 diabetes and poor glycemic control have been identified as common factors influencing the incidence of heart failure. In people with diabetes, a significantly increased risk of heart failure has been demonstrated [38]. Additionally, annual mortality of diabetic patients with heart failure is increased by more than 10-fold as compared with diabetic patients without heart failure [38].

Echocardiography has been established as the preferred method for documentation of cardiac dysfunction. The left ventricular (LV) ejection fraction (EF) is the most important measurement for the assessment of impaired systolic function. In patients with normal or mildly abnormal LV systolic function, abnormal LV relaxation, diastolic distensibility, or diastolic stiffness are echocardiographic signs of diastolic dysfunction. Echocardiography, including tissue Doppler imaging is useful in detecting myocardial dysfunction in diabetic patients, as well as in nondiabetics [4].

Measurement of plasma concentrations of natriuretic peptides or their precursors for diagnosing heart failure may also be helpful in patients with diabetes.

Diastolic dysfunction is a well-established abnormality in diabetes mellitus, probably in the context of hypertension and perhaps cardiomyopathy leading to a disordered filling of the left ventricle. Quantitative and qualitative changes of the extracellular matrix formation of the heart have been observed dependent upon hyperglycemia and an excess of glucose in animal and in vitro studies and may play a role also in the human situation [28].

Echocardiographic investigations show a rapid improvement of heart muscle performance with stimulating glucose uptake by insulin administration in diabetic patients with ischemic heart disease. On the contrary, abundant concentrations of free FA may weaken cardiac pump function and, in addition, induce severe arrhythmias.

Blood flow and in particular reactive hyperemia induced by hypoxia may be impaired even in the absence of advanced atherosclerosis in view of profound endothelial dysfunction, and also activation, both of which seem to occur quite early on in diabetes [28].

Microvascular disease of the heart
Microangiopathy represents a common and widespread manifestation in patients with multiorgan involvement. Abnormalities of coronary microcirculation have been reported in clinical diabetes mellitus using invasive or noninvasive procedures. Noninvasive studies using positron emission tomography (PET) with N-13 ammonia demonstrated a largely unaltered maximal endothelial-independent myocardial vasodilatory capacity induced by adenosine or dipyridamole stress testing and an impaired response to sympathetic mediated stimulation by cold pressor test (CPT) in diabetic patients, which assessed endothelial-dependent blood flow [39].

Dysregulation of coronary vascular function in response to CPT was found to be present in approximately one third of those diabetic patients without clinical signs of coronary artery disease [39]. Diabetic patients presented with higher baseline myocardial blood flow (MBF) at resting conditions and a paradoxical coronary vasoconstriction during the CPT [39]. Scintigraphic assessment, therefore, demonstrates complex patterns of dysregulation of myocardial blood flow in diabetes, which is even present in the absence of coronary artery disease.

Arrhythmias
Due to reduced cardiac ejection fraction and systolic dysfunction, impaired sympathetic and parasympathetic responses, and decreased diastolic filling, CAN is associated with a reduction in exercise tolerance [40]. An imbalance between right and left sympathetic innervation is reported to correspond with a prolonged corrected QT interval and QT dispersion (the difference between the longest and shortest QT interval) [40]. Diabetic patients with a regional sympathetic imbalance and QT interval prolongation are suggested to be at higher risk for arrhythmias [40]. Hypoglycemia has been found to contribute significantly to increased risk of arrhythmias (see earlier and following section).

Cardiovascular risk related to diabetes therapy
Defining the vulnerable patient with critical hypoglycemia
Data from long-term controlled interventional trials with intensified glycemic control such as the DCCT for T1DM [41] and the UKPDS for newly diagnosed T2DM [42] have shown fewer cardiovascular events with better glucose control. Similarly, the incidence of major cardiovascular events was reduced by 17% with more intensive diabetes control in a recently published meta-analysis of five mega-trials; however, no effect was seen for all-cause mortality or stroke [20]. Intensified glucose control with basal insulin glargine in the ORIGIN trial, however, had no effect on any cardiovascular event rate compared to standard care with metformin and sulfonylureas when necessary, despite a sustained maintenance of HbA1c <6.5% [19]. In all these studies (see also Table 73.1) intensified glucose-lowering treatment was associated with an increase in hypoglycemic episodes. Furthermore, severe hypoglycemia, defined as any episode that requires external assistance, was associated with an increased mortality rate in all these studies [43]. In the DIGAMI 2 study involving patients with acute coronary syndrome, long-term insulin treatment was associated with significantly higher mortality compared to active comparator metformin; this notion seems to point to the fact that a strategy of antidiabetic treatment with higher risk of hypoglycemia may trigger cardiovascular complications [44]. However, in patients hospitalized with acute myocardial infarction, iatrogenic hypoglycemia after
insulin therapy was not associated with increased mortality [45]. Further of note, mild hypoglycemia, for example self-reported and self-treated episodes, was not associated with a significant increase in mortality in a large retrospective trial using the Charlson comorbidity index to minimize selection bias by the decision of more intensive glucose-lowering treatment [46]. In a large retrospective registry study with 845 type 2 diabetes patients, 3.1% reported a hypoglycemic event. The odds ratio for acute cardiovascular events for those with hypoglycemia was 1.79 [47]. So far, little is known about the contribution of harmful cardiovascular effects of hypoglycemic episodes such as incident arrhythmias, ischemic reactions, and increased cardiac work load on major CV events and mortality. Investigations with continuous glucose measurement systems (CGMS) show a close relationship between episodes of critical low glucose levels (below 3.1 mmol L⁻¹) and ventricular arrhythmias as well as ischemic changes in parallel ECG recording [48,49]. Rapid glycemic fluctuations produce oxidative stress [50] with potential harmful effects on electrical stability of the heart. Thus, impaired glucose homeostasis could precipitate major CV events in a multistage metabolic and hormonal stress reaction leading to electric instability and ischemic reactions in patients with acute coronary syndrome or multivessel disease [51].

**Mechanisms and links to cardiovascular complications during critically low glycemia**

Sudden death in patients with T1DM without coronary heart disease observed during sleep has been described as “dead-in-bed” syndrome [52]. Parallel measurements with Holter ECG and CGMS revealed a close association of long-lasting severe hypoglycemia triggering prolongation of QTc time and ending up with fatal tachyarrhythmias [53]. However, dead-in-bed syndrome is rare in young adults with T1DM without CHD despite a high frequency of nocturnal hypoglycemia [54]. The link between severe long-lasting hypoglycemas and fatal ventricular arrhythmias, however, illustrates that hypoglycemia may trigger serious CV complications in frail type 2 patients.

Hypoglycemia has long been known to be associated with hemodynamic alterations such as increase in heart rate, ectopic heart beats, chest pain, and increase in systolic blood pressure as a consequence of activation of the sympatho-adrenal system. Physiologic responses are release of stress hormones, in particular catecholamines, ACTH/glucocorticoids, and glucagon to protect end-organs and increase gluconeogenesis. At critical low glucose levels, however, this leads to an increased workload of the heart, impairs endothelial function and may have harmful effects on electric stability in vulnerable patients with CHD [54]. Mechanistic studies applying iatrogenic hypoglycemia have shown that a cascade of sympatho-adrenal counterregulations starts to stabilize glucose supply to end-organs, in particular in the brain, at levels of plasma glucose <3.9 mmol L⁻¹ [51]. However in healthy subjects, long-lasting low glucose levels (below ~3 mmol L⁻¹) cause life-threatening ECG changes with lengthening of the QT interval. This prolongation of cardiac repolarization predisposes to cardiac arrhythmias. Stress hormones such as catecholamines also lower serum potassium, a key player for electrophysiologic stability, which may further increase risk of arrhythmias [55]. By using radionuclide ventriculography it could be shown that hemodynamic changes due to insulin-induced hypoglycemia produced a sustained effect also on cardiac contractility which may be explained by prolonged and impaired repolarization during diastole [56]. Besides these hormonal/metabolic effects, there are other secondary critical changes that may affect thrombus formation and plaque stability in patients with CHD: activation of platelets and coagulation and increase in inflammatory cytokines [57].

Another critical aspect associated with hypoglycemia is its effect on autonomic cardiovascular function. As proven by CLAMP-studies in healthy subjects, baro-reflex sensitivity and appropriate sympathetic-vagal response to hypotensive stress are attenuated subsequent to hypoglycemia [58].

**Coincidence of hypoglycemia and arrhythmias**

Atrial fibrillation (AF) is a frequent finding in patients with T2DM. In the Atherosclerosis Risk in Communities Study the hazard ratio for AF in patients with T2DM was 1.35. However, prediabetes and newly diagnosed diabetes was not associated with increased risk of AF compared to subjects without diabetes [59]. In multivariate analysis diabetes, HbA1c and poor glycemic control were independently related to the risk of incident AF. Atrial fibrillation is common in diabetic patients with acute myocardial infarction. In a Russian study of patients with myocardial infarction, 42.1% with T2DM had extra-systolic beats and AF versus 30.4% without diabetes. Cardiac arrhythmias were directly correlated with coronary and myocardial dysfunction and were prognostically unfavorable factors for survival rate [60]. Furthermore, in patients with heart failure rapid changes in glucose levels measured with CGMS as mean amplitude of glucose excursions (MAGE) were associated with serious arrhythmias at MAGE >5 mmol L⁻¹.

As a consequence, a significant increase in cardiac mortality was observed in this very high-risk population [61]. No benefit of aggressive glucose control could be observed in patients with diabetes undergoing coronary artery bypass graft surgery because of a higher incidence of hypoglycemic episodes with harmful effects on cardiac work load and electric stability [62]. Parallel recording of CGMS and ECG in vulnerable patients with advanced T2DM and CVD treated with insulin or/sulfonylurea reveals a high incidence of asymptomatic nocturnal episodes of severe hypoglycemia associated with silent ventricular tachycardias (Figure 73.1)[63]. Cardiac autonomic neuropathy has been identified to be associated with an excessive mortality ([64], see also earlier in this chapter). It aggravates risk of ventricular tachycardia in the case of hypoglycemia [65]. By extrapolation of these data in accordance with the anecdotal finding of dead-in-bed syndrome in T1DM, it may be suggested that severe hypoglycemia via arrhythmias and coronary ischemia fuels a vicious cycle which may cause
Figure 73.1 High fluctuations are associated with arrhythmias.

Male, 56 years, BMI 43.3, HbA1c 6.9%, previous apoplexia
Therapy: Metformin 2000 mg, Detemir 60 IU, Humalog 70 IU
sudden death by ventricular fibrillation or heart failure in frail patients. However, there is no evidence at present for causality between overall hypoglycemia and cardiovascular events. Outside of cardiovascular clinics with continuous monitoring of ECG and glucose levels only a few studies with 2–5 days of ECG recordings exist. Thus, the contribution of arrhythmias induced by hypoglycemia to an adverse outcome in patients with CVD or with heart failure under real-world conditions is uncertain.

**Impact of antidiabetic treatment on risk of hypoglycemia and arrhythmias**

The risk of cardiovascular events associated with hypoglycemia related to antidiabetic drugs has been a matter of controversial debate for a long time. Only scarce data is available for insulin and sulfonylureas. Both classes of drugs can cause long-lasting hypoglycemia and rapid glucose fluctuations with possible effects on myocardial blood flow and electric stability. In the UKPDS legacy analysis, intensified treatment with sulfonylurea and/or insulin was associated with significantly lower cardiovascular event rate and all-cause mortality compared to conventional therapy, despite a higher rate of hypoglycemic episodes during the intervention phase with those classes of drugs [46]. In the ADVANCE study, intensified treatment with the sulfonylurea gliclazide was associated with increased risk of hypoglycemia, but no association between hypoglycemia and cardiovascular mortality was observed [43]. Mechanistic studies, however, suggest that sulfonylurea drugs may have harmful effects on myocardial ischemia tolerance and electric stability in patients with acute coronary syndrome. In a comparative trial of patients with acute myocardial infarction undergoing angioplasty in-hospital mortality was significantly higher in patients treated with sulfonylurea (OR 2.77). Sulfonylurea intake at time of acute myocardial infarction, however, was not related to late outcome [67]. In a CLAMP study, heart rate variability in response to hypoglycemia was impaired in patients with T2DM after a single dose of glibenclamide in the morning, resulting in a higher risk for critical arrhythmias during hypoglycemic episodes [68]. Strict glucose control with basal insulin glargine versus standard care in the ORIGIN trial was associated with a significantly higher incidence of severe hypoglycemic episode (1.0% vs. 0.3% per year), but cardiovascular outcome was neutral and only one death may have been related to hypoglycemia [19]. Thus, data derived from DIGAMI 2 [44] and the European Heart Survey [69] suggesting a higher risk with insulin treatment could not be confirmed. Furthermore, a large retrospective cohort of patients hospitalized with acute myocardial infarction in 40 hospitals in the US was evaluated for the relationship between spontaneous and insulin-induced hypoglycemia and mortality. Iatrogenic hypoglycemia after insulin treatment was not associated with higher mortality, while spontaneous hypoglycemia showed an increased OR for mortality of 2.32 [45]. This puzzling finding may be explained by the fact that insulin counteracts oxidative stress due to rapid glycemic fluctuations as shown in CGMS-based cohort investigations [70].

By extrapolation so far, there is no definite evidence for additional harmful effects of insulin or sulfonylureas in the context of hypoglycemic episodes in long-term studies. In acute coronary syndrome sulfonylurea may deteriorate tolerance to ischemia.

**Conclusion**

Review from the literature, mechanistic studies and clinical experience in patients with CHD and heart failure suggest that severe hypoglycemia and rapid glucose fluctuations can initiate a cascade of pro-arrhythmic and ischemic reactions which may lead to fatal cardiovascular events or critical deterioration of heart failure. Pre-existing CHD and autonomic neuropathy label vulnerable patients. Antidiabetic drugs are the most common cause of hypoglycemia. No specific effects of glucose-lowering drugs on risk of hypoglycemia are known so far. While we have only proof for causality from in-hospital patients with acute coronary syndrome and some anecdotal cases with dead-in-bed syndrome, intensive glucose-lowering therapy in frail and elderly patients that could provoke critical low glucose levels and rapid fluctuations should be avoided.

**Management aspects**

**No clear evidence for routine screening, but for risk reduction**

The Detection of Ischemia in Asymptomatic Diabetics (DIAD) study, a randomized controlled trial in which 1123 participants with T2DM and no symptoms of coronary artery disease (CAD) were randomly assigned to be screened with adenosine-stress radionuclide myocardial perfusion imaging (MPI) or not to be screened, aimed to assess whether routine screening for CAD identifies patients with T2DM as being at high cardiac risk and whether it affects their cardiac outcomes [71]. The incidence of cardiovascular events was low in this cohort and thus the DIAD study failed to prove the benefit of a generalized screening of patients with diabetes mellitus. However, the ADDITION trial in which new diagnosed diabetic patients with no previous cardiovascular event were included and followed for 14 months showed a downward shift in the distribution of modeled cardiovascular disease (CVD) risk over 14 months, using the UKPDS risk engine. Older individuals, males, and those with a larger waist circumference at baseline exhibited smaller risk reductions. Individuals prescribed a higher number of drugs over the follow-up period, and those who decreased their energy intake or reduced their weight, demonstrated larger reductions in the modeled risk. This study revealed a beneficial effect of metabolic screening and of early intervention, and demonstrates the power of individual risk scoring following diagnosis of diabetes by using the UKPDS risk score in terms of reduction of cardiovascular risk [72].
The diabetic patient after the cardiovascular event
Up to one quarter of patients referred for coronary revascularization present with diabetes mellitus [73,74]; of those one third are admitted with acute coronary syndromes (ACS), and more than one third of patients present with cardiogenic shock. Diabetic patients undergoing revascularization procedures show a higher prevalence of previous myocardial infarction, congestive heart failure and arterial hypertension compared to nondiabetic patients [75] and the clinical outcome is impaired. Coronary artery disease is more severe in diabetic than in nondiabetic patients, the prevalence of three-vessel disease is more than 50% [76]. Heart disease in patients with diabetes has to be regarded as a composite of hypoxia through ischemia and malnutrition through impaired myocardial metabolism.

Furthermore, diabetes is associated with higher rates of restenosis after PCI of each type (balloon angioplasty or implantation of bare metal and drug-eluting stent) and is regarded as an independent predictor of early stent thrombosis both in the bare metal (BMS) and drug-eluting stents (DES) [77]. DES are more effective than BMS in reducing the rate of repeat intervention and MACE, although they have shown similar rates of death and myocardial infarction [78,79]. Although medical treatment has improved significantly over recent years the long-term risk of death, myocardial infarction and repeat revascularization is nearly twice as high in diabetic than in nondiabetic patients undergoing PCI [80]; ACC/AHA [81] and EAC/EACTS [82] guidelines recommend coronary artery bypass grafting (CABG) rather than PCI for diabetic patients with stable angina and multivessel coronary disease.

Lessons learned from revascularization trials
The main finding of the Bypass Angioplasty Revascularization Investigation (BARI) trial performed around 17 years ago was that patients with diabetes and multivessel disease undergoing CABG had a better prognosis for survival than patients in whom balloon angioplasty has been performed [83]. As a result the National Heart, Lung, and Blood Institute issued a clinical alert and CABG was recommended as the preferred approach for revascularization in these patients [84,85].

BARI findings were subsequently underlined by several studies showing that more major adverse cardiovascular and cerebrovascular events in patients with diabetes and multivessel CAD occurred who underwent PCI than among those in whom CABG was performed [86]. Significant improvements in antiplatelet therapy with availability of highly potent drugs (ADP receptor agonists and glycoprotein IIb/IIIa antagonists) mainly contributed to the higher success rates for CABG in multivessel disease. In particular, the diabetic patient profits from post interventional treatment with ADP receptor agonists [87,88].

Even the comparison of CABG to drug-eluting stents performed in the Coronary Artery Revascularization in Diabetes (CARDia) study [89] or by analyzing the diabetic subgroup of the Synergy between PCI with TAXUS and Cardiac Surgery (SYNTAX) study clearly showed the benefit of CABG. The SYNTAX trial compared 1800 patients with symptomatic left main and/or three-vessel coronary artery disease (CAD) treated with drug-eluting stents versus coronary artery bypass graft surgery [90]. Major adverse cardiovascular and cerebrovascular events at 12 months were more often for patients with multivessel disease who underwent PCI with drug-eluting stents compared to CABG [89,90], an effect which was even more prominent in the long-term follow-up of the SYNTAX patients [91].

A subgroup analysis from the COURAGE (Clinical Outcomes Utilizing Revascularization and Aggressive Drug Evaluation) trial furthermore suggested that the addition of early PCI to optimal medical therapy did not reduce the rate of adverse events after a median follow-up of 4.6 years [92]. The COURAGE trial randomized 2287 patients with stable angina and CAD to PCI plus optimal medical therapy versus optimal medical therapy alone, with a primary endpoint of all-cause mortality and nonfatal myocardial infarction (MI) [93]. In all subgroups, the outcomes of death or MI were similar in patients who underwent PCI or those who did not. In BARI 2 Diabetes (BARI 2D) this estimation was confirmed for diabetic patients as well [94].

The Future Revascularization Evaluation in Patients with Diabetes Mellitus: Optimal Management of Multivessel Disease (FREEDOM) trial is the most recently published study on this topic [95]. CABG was analyzed for superiority to PCI with drug-eluting stents in the treatment of 1900 high-risk patients with a good distribution of SYNTAX score and diabetes (about as many patients with diabetes as in all previous trials combined) and multivessel coronary artery disease. Consistent with the BARI 2D trial, optimal medical therapy was prescribed throughout the follow-up. After long-term follow-up (median 3.8 years), 947 patients assigned to CABG showed significantly lower mortality rate (10.9% vs. 16.3%) and fewer myocardial infarctions (6.0% vs. 13.9%) than 953 patients assigned to undergo PCI. However, patients in the CABG group presented with significantly more strokes (5.2% vs. 2.4%), mostly because of strokes that occurred within 30 days after revascularization and thus the benefit of CABG was driven by reductions in rates of both myocardial infarction and death from any cause. Regarding the primary composite outcome of death from any cause (myocardial infarction or stroke over 5 years), the incidence was reduced in the CABG group by 7.9 percentage points (18.7% vs. 26.6%; p = 0.005), as compared to PCI [95]. Taken together the results from 13 trials in more than 4000 patients with diabetes reveal that mortality has been consistently reduced by CABG, as compared with PCI of each type [91].

The metabolic component
Regarding the metabolic component in the elective treatment of the vascular diabetic patient no clear evidence for a benefit of tight HbA1c adjustment (HbA1c <7.0%) is available.
However, a BMI of $\leq 30 \text{ kg m}^{-2}$ turned out to be of advantage in terms of outcome, a fact which is known as obesity paradox [96]. The meta-analyses of prospective randomized clinical trials including ADVANCE, ACCORD, and VADT clearly point to the reduction of fatal and nonfatal infarction and the risk of coronary events with intensified glucose control in terms of prevention. For an optimized therapeutic result the individual patient and his comorbidities have to be taken into account. Regarding the most efficient therapy, side effects like hypoglycemia should be considered and avoided. A multifaceted therapeutic concept which includes lifestyle modification (smoking cessation, dietary concepts, and physical exercise) should be combined with antithrombotic, lipid lowering, antihypertensive, and blood glucose normalization strategies. Strong evidence for this concept with a number needed to treat (NNT) below 10 is based on the Steno-2 trial which took into account high-risk patients and showed significant reduction of morbidity and mortality in these patients [97,98]. In combination with commonly used risk scores like UKPDS or CARRISMA patients at risk can be properly evaluated and treated [99,100]. It becomes obvious that not one factor per se but many factors in concert determine the outcome of patients with diabetes and heart disease. This again points to the fact that diabetic heart disease is not only an ischemia-driven disease but is also characterized by a strong metabolic component [100,101].

Concluding remarks

Not only do the recent published studies prove the superiority of CABG over PCI in the treatment of multivessel disease but newer developments in surgical techniques, such as minimal invasive procedures and “off-pump” surgery, which decrease manipulation of the ascending aorta with the risk of atherosclerotic emboli and activation of the complement system, potentially leading to multiple organ dysfunction and/or damage, will give additional substance to the preference of CABG over PCI. This is of particular interest in diabetic patients with more extensive atherosclerosis and higher risk of perioperative infection.

References

17. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent


Clinical features and treatment of coronary heart disease in diabetes 1077


CHAPTER 74
Arterial hypertension in diabetes: etiology and treatment

Colleen Flynn and George Bakris
Department of Medicine, Section of Endocrinology, Diabetes and Metabolism and the ASH Comprehensive Hypertension Center, The University of Chicago, Chicago, IL, USA

Key points
- Hypertension is a polygenic disorder that commonly accompanies type 2 diabetes and contributes to decline in kidney function as well as cardiovascular risk.
- Increased sympathetic tone is a key factor in driving elevations in blood pressure in patients with diabetes.
- Endothelial dysfunction associated with salt sensitivity increases blood pressure variability and contributes to increased risk of stroke.
- Factors early in diabetes that improve endothelial function and reduce insulin resistance such as exercise and weight loss are key to reducing the risk of blood pressure elevation.
- Randomized clinical trials support blood pressure goals below 140/90 mmHg but not <130/80 mmHg for cardiovascular risk reduction.

Introduction
Hypertension, defined as a consistent systolic blood pressure above 140 mmHg or a diastolic blood pressure above 90 mmHg on two separate occasions measured using American Heart Association criteria [1], affects more than 72 million Americans and is one of most prevalent risk factors that contributes to development of cardiovascular disease and chronic kidney disease [2,3]. The prevalence of hypertension is estimated at about 30% of the adult population in developed countries and is predicted to increase by almost 60% in the next two decades [4]. Diabetes is a major risk factor for cardiovascular disease [2]. Cardiovascular morbidity and mortality is increased substantially in the presence of diabetes [2,5]. In the natural history of type 1 diabetes (T1DM), development of an elevated BP (i.e., >130/80 mmHg) is a major predictor of nephropathy and future decline in kidney function; this is especially true in those with a progressive rise in blood pressure over time and especially in those with a family history of kidney disease [2,6]. In contrast, hypertension is already evident in most patients with T2DM at the time of diagnosis. The implications of hypertension on cardiovascular risk, however, are similar in both types of diabetes [2,7]. Mortality is increased 7.2-fold when hypertension is present in patients with diabetes [2].

The contributing factors in the pathogenesis of hypertension in diabetes are multifactorial. For many years, it has been recognized that hypertension is common to both obese subjects and those with T2DM [8]. Blood pressure elevations in both these groups may be, in part, due to the presence of insulin resistance and resultant hyperinsulinemia as well as increased salt sensitivity. Support for this proposal is evident from studies of behaviors that improve insulin action, such as weight loss and increased physical activity and the resultant reduction in blood pressure into the normal range [9–11]. Additionally studies that are more recent demonstrate blunting of vasodilator responses to known stimuli and reduced nitric oxide release; factors associated with increases in salt sensitivity, that is, both systolic and diastolic blood pressure increase by >5 mmHg more than someone given the same salt load of 400 mmol d⁻¹ [12,13].

To understand the contribution of insulin resistance in the genesis of hypertension fully, one has to evaluate the effects of insulin resistance and hyperinsulinemia on factors that contribute to blood pressure elevation. Insulin resistance is a metabolic disorder that is manifested by a reduction in peripheral skeletal muscle utilization of glucose, fatty acid, and protein metabolism [14]. Hence, higher concentrations of insulin (hyperinsulinemia) are needed to achieve the same level of glucose utilization in these tissues. Hyperinsulinemia is also associated with a number of physiologic changes in cellular function. Hyperinsulinemia may also contribute to the genesis of hypertension through its effects on sodium homeostasis and the sympathetic nervous system. Lastly, the effect of insulin
on various growth factors contributes to the development of vascular injury through its potentiation of the atherosclerotic process [15].

As previously noted, most obese individuals develop insulin resistance and subsequent hyperinsulinemia. Yet, they do not all develop T2DM or hypertension as defined by the most recent Joint National Committee (JNC 7) Report [16]. The reasons for this lack of consistency with disease development are likely due to varied genetic and environmental factors that will be discussed.

**Genetics factors**

Work by various investigators to isolate a “hypertensive gene” or group of genes has been ongoing for many years [17]. Many candidate genes for hypertension have been found, including genes for ion channels within the kidney that affect transporters (SLC12A3, SLC12A1, KCNJ1, SCNN1A, SCNN1B, SCNN1G, CLCNKB), ion channel regulation (WNK1, WNK4, SGK1, ADD1, ADD2, GRK4), aldosterone signaling (REN, AGT, ACE, AGTR1), catecholamine pathways (TH, COMT, DBH, DRD1, DRD2, ADRB1, ADRB2, ADRB3, ADRA1A), vasoconstriction (NOS3, EDN1, EDNRA, CYP2C8), and inflammation (TGF-beta) [18].

There are also studies investigating for genes that increase risk for development of both hypertension and diabetes. Kraja et al. evaluated the 22,161 participants in the seven studies of the STAMPEED consortium to determine if subjects with the metabolic syndrome phenotype had common genetic variants. Using genome-wide association analyses, they identified two single-nucleotide polymorphisms (SNPs) located between LOC100128354 (similar to small nuclear ribonucleoprotein single-nucleotide polymorphisms (SNPs) located between LOC100128354) and angiotensin type 1 receptor gene polymorphism is associated with hypertension [22].

A study by Patel et al. investigated variations in the ACE2 (angiotensin-converting enzyme 2) gene in Caucasian men and women with T2DM. ACE2 is a homologue of ACE, and degrades angiotensin II to the vasodilator angiotensin I-7. The researchers found certain polymorphisms of ACE2 in men and women with T2DM that were associated with hypertension [20]. Bengtsson et al. investigated polymorphisms of the ACE gene and angiotensinogen (AGT) gene and their association with hypertension and T2DM. They found the D-allele of the ACE gene ID polymorphism increases susceptibility to hypertension, particularly when associated with T2DM [21]. A smaller study of 69 subjects with insulin-dependent diabetes found the CC-genotype of the A1166C gene polymorphism of the angiotensin type 1 receptor gene polymorphism is associated with hypertension [22].

A number of other factors have been associated with hypertension development in diabetes. Haptoglobin polymorphisms were examined in a 120-subject study, the results of which predicted the development of hypertension in patients with diabetes. The Hp1-2 genotype was the most common among those with refractory hypertension and T2DM [23]. Renalase is an important enzyme in catecholamine metabolism and is linked with changes associated with worsening of cardiac ischemia and kidney disease [24–26]. Buraczynska et al. found an association between the renalase gene polymorphism (C allele of rs2296545 SNP) with hypertension in 892 patients with T2DM [26]. Lastly, a recent study by Groop et al. investigated the association between polymorphism of the glycogen synthase gene and the associated risk of developing T2DM [27]. These investigators found two polymorphic alleles (A1 & A2) in this gene located on chromosome 19. Moreover, they documented a twofold greater prevalence of hypertension among subjects without diabetes who had the A2 allele expressed. Unfortunately, no such distinction was noted in the group studied with diabetes. Note, however, that among the group with diabetes, both the A1 & A2 allelic groups had hyperinsulinemia and were hypertensive. Taken together, these studies provide evidence that supports the concept that a different “genetic load” is required for the development of either hypertension or diabetes. Specifically, each of these disorders appears to be inherited in a polygenic (i.e., to involve more than one gene) and heterogeneous (involving different constellations of disease genes in different persons) fashion. Since hypertension does not develop in all people with diabetes, presence of certain environmental factors, that is, diet, sedentary lifestyle, high salt intake, etc., are mostly required to express the concomitant generation of both diseases.

**Ion transport homeostasis**

Substantial evidence from animal models of hypertension as well as diabetic and nondiabetic hypertensive individuals supports an association between the presence of hypertension and changes in intracellular pH as well as electrolyte composition [17,28–36]. These observations have led to various hypotheses regarding the relative importance of one ion over another. This section reviews the studies for each ion implicated in causing hypertension as well as the effects of intracellular pH change.

Numerous investigators have documented increases in cytosolic free sodium concentrations from cells of hypertensive or diabetic patients when compared to age- and sex-matched nondiabetic controls [30,34,37,38]. These increases result from altered activity of the Na/H antipporter and the Na/Li countertransporter. The increases in intracellular sodium are highly correlated with the presence of an elevated diastolic blood pressure. Conversely, no relationship exists between increases in intracellular sodium concentration and the development of microalbuminuria [33].
The relationship between intracellular magnesium and blood pressure is less clearly defined. Hyperglycemia is known to reduce intracellular magnesium [36]. Moreover, data from experimental models of hypertension as well as from diabetic subjects with hypertension demonstrate an inverse relationship between intracellular magnesium concentration and blood pressure elevation [38]. The primary mechanism responsible for this relative reduction in intracellular magnesium relates to sodium-dependent magnesium efflux through the plasmalemmal membrane [38].

Studies in experimental models of diabetes consistently demonstrate abnormally elevated concentrations of intracellular calcium [28]. Moreover, increases in the intracellular calcium concentration are commonly seen in diabetic hypertensive subjects as well as in obese and essential hypertensive subjects [30]. Different combinations of circulating insulin and glucose levels in diabetes, obesity and essential hypertension may help explain observed differences in calcium regulation [30]. Hyperglycemia increases intracellular calcium concentration [36]. Since intracellular calcium has an optimal range for mediating insulin action, these changes result in an impaired cellular ability to generate acute calcium signals [36]. This consequently attenuates insulin action and thus, impairs glucose transport, decreases insulin-stimulated glucose uptake and ultimately contributes to the genesis of insulin resistance [39]. Lastly, there is evidence that the density of calcium channels is increased in peripheral skeletal muscle of diabetics, which may relate to these intracellular changes in calcium [28]. Moreover, there may be an interaction with regulation of calcium channels and the insulin receptor gene that remains to be elucidated. An in-depth analysis of studies relating to altered intracellular calcium homeostasis in diabetes is beyond the scope of this chapter. Taken together, these studies suggest that the concentration dependent influx of calcium into cells may be an initiating event that contributes to the pathophysiology of hypertension in diabetes. This is indirectly supported by early studies showing that moderate dose nifedipine failed to lower blood pressure in normotensive individuals but will lower pressure in hypertensive people [40,41].

The key cellular changes, therefore, in individuals with insulin resistance and hypertension may be summarized as an increase in the intracellular sodium and calcium and a decrease in magnesium composition [30,36,38,39,42–46]. Additionally, reductions in intracellular pH are also well documented from in situ studies of cells from both animal models of hypertension as well as hypertensive subjects [39,47,48]. The observed reductions in intracellular pH come from in situ studies and are contrary to other reports using tissue culture techniques. These latter studies report an increase in intracellular pH in subjects with diabetes [48]. This discrepancy may be due to the differences in pH measurement using a tissue culture technique versus in situ measurements. In situ studies, however, generally require minimal processing and hence are a better reflection of the disease state in human beings. Interestingly, glucose ingestion increases pH in normotensive nondiabetic subjects [36,39].

The aforementioned changes in cellular milieu are linked to altered function of membrane ion transporters (Na/H antipporter, Na/K/ATPase, Mg/Na exchanger, Ca/H exchanger, Ca ATPase, and others). Both the sodium potassium ATPase and the Ca ATPase pumps are important in maintaining calcium homeostasis of the cell. Increased calcium ATPase activity has been reported in the renal cortex of non-insulin-dependent diabetic rats and in the platelets of non-insulin-dependent diabetic patients [49]. The converse is true, however, in insulin-dependent diabetic subjects [30]. These differences may relate to potentially different physiologic mechanisms that generate elevations in blood pressure.

Hormonal regulation of intracellular calcium may be abnormal in diabetes. Insulin and atrial natriuretic peptide lose their ability to stimulate the activity of calcium ATPase pumps in the renal cortex of non-insulin-dependent diabetic rats. Moreover, these hormones catalyze an exaggerated increase in platelet intracellular calcium in the presence of thrombin, adenosine diphosphate, and low-density lipoproteins in patients with both T1 and T2DM. Lastly, studies in hypertensive animal models as well as in patients who have hypertension and T2DM support the concept that these cellular changes are predominantly seen in salt-sensitive forms of hypertension [33,39,43,47,48].

It is difficult to assess whether alterations in these ion transporters antedates the development of hypertension or diabetes or whether they are a consequence of metabolic changes that subsequently generate a cellular milieu that leads to hypertension. One proposed hypothesis to explain these ion changes is that a primary (genetic) mutation is present on the gene locus that controls the Na/H antiporter and that this defect is expressed in the presence of persistent hyperglycemia [29,36,50]. This environmentally induced gene expression then alters cellular ion and pH homeostasis and directly contributes to the genesis of the insulin-resistant syndrome, T2DM, hypertension, and lipid abnormalities [29,50].

A number of recent studies support the hypothesis that cellular ion changes antedate the development of hypertension in diabetes [35,51,52]. However, only one of these studies provides direct evidence that hyperglycemia serves as the initiating event to alter Na/H antiporter activity [52]. Further support that Na/H antiporter activity is increased in diabetic hypertensive subjects comes from clinical studies that document increased activity of this pump in both nondiabetic and diabetic hypertensive subjects [28,31,53]. Genetic studies using linkage analysis, however, have failed to provide any evidence that mutations of the Na/H antiporter gene directly contribute to the pathogenesis of hypertension [54,55]. Similar disappointing results were noted for the Na/Li countertransporter [54]. Therefore, until evidence that is more conclusive is available the role of these ion transporters directly linking to the pathogenesis of
hypertension in diabetes is questionable. The development of hypertension in patients with T1DM is primarily related to the progression of nephropathy [56]. In contrast, the mechanism of hypertension in T2DM is primarily related to insulin resistance and central obesity [57].

**Extracellular sodium homeostasis**

Under normal conditions, various stimuli that raise blood pressure also increase sodium and water excretion through pressure natriuresis and diuresis. This natriuretic response continues until the reduction in venous return and cardiac output attenuate increases in blood pressure and fluid balance is restored. A number of hormonal and hemodynamic factors influence sodium excretion in people who have hyperinsulinemia, that is, obese or diabetic subjects. Hyperinsulinemic individuals are both salt sensitive and have an increase in total body sodium [58].

Numerous studies demonstrate a direct link between hyperinsulinemia and increased tubular sodium reabsorption. DeFronzo et al. were the first to report that at high physiologic insulin concentrations in the presence of euglycemia healthy humans reduced urinary sodium excretion by about 50% [29]. This observation is further corroborated by more recent studies with insulin concentrations of <70 μU mL⁻¹ [59]. It is noteworthy that plasma insulin levels of 35 μU mL⁻¹ can elicit a mild antinatriuretic effect, a level noted in fasting insulin concentrations among obese individuals [60]. This antinatriuretic effect of insulin is related to an increased sodium reabsorption present primarily in the distal and to a lesser extent in the proximal tubule [61,62].

Chronic hyperinsulinemia serves as a better model to study the effects of insulin on sodium balance since it permits time for the kidney to “escape” the sodium-retaining effects of insulin. Other studies utilizing sodium loading of uncomplicated diabetic hypertensive subjects failed to demonstrate a difference in natriuretic response between diabetic and normal control subjects. Studies by Hall et al., following one month of chronic insulin infusion in dogs with reduced renal mass fed a high-salt diet failed to demonstrate a reduction in sodium excretion or an increase in blood pressure [60,63]. While hyperinsulinemic, these dogs were not hyperglycemic; however, insulin resistance is associated with higher levels of insulin which at the level of the proximal and distal tubule enhance sodium reabsorption, and hence it contributes to increases in blood pressure [64].

Chronic hyperglycemia is consistently associated with an increase in exchangeable body sodium [65]. Contrary, to hyperinsulinemic euglycemic models, hyperglycemic dogs in this study demonstrated a reduced sodium excretion in response to saline loading [65]. Taken together, it appears that hyperglycemia through effects on the sodium glucose exchanger in either the proximal tubule or other yet unknown mechanism increases total body sodium. Moreover, pressure natriuresis effectively balances insulin-mediated sodium retention under conditions of chronic hyperinsulinemia. Thus, hyperinsulinemia alone is not expected to be invariably associated with an expansion of the body sodium pool but rather hyperglycemia mediates this effect and hence, hyperinsulinemia alone does not contribute to the genesis of hypertension in this way.

Multiple mechanisms are postulated to explain the antinatriuretic response of insulin. One such mechanism is the direct inhibitory effect of insulin on atrial natriuretic peptide (ANP). Numerous studies in both animal models of diabetes as well as in humans demonstrate an increased level of ANP as well as inhibition of its effect by insulin [66,67]. Chronic infusion of insulin results in a transient reduction in sodium excretion from which the kidney “escapes” [62]. The antinatriuretic response to insulin in normotensive subjects is secondary to increased distal tubular sodium reabsorption [62]. This response to insulin, however, is inhibited by ANP through its direct action on the inner medullary collecting duct; ANP does not inhibit this response in hypertensive subjects [68,69]. Thus, in hypertensive subjects the resistance to ANP appears to be localized in the inner medullary collecting duct.

Chronic hyperglycemia is consistently associated with an average 10% increase in body sodium [70]. Moreover, sodium excretion and response to saline loading or water emerging is blunted in such patients. This may relate to decreased effectiveness of natriuretic hormones such as ANP. However, in non-azotemic subjects with diabetes, given normal circulatory blood volumes, sodium concentrations are normal. This coupled with the hyperosmolarity-induced hyperglycemia and reduction in oncotic pressure suggests that chronic hyperglycemia is associated with an expanded extravascular fluid volume at the expense of intracellular and intravascular fluid volumes. Thus, under these circumstances, pressure natriuresis is no longer capable of balancing out the excess sodium retained by the kidneys. Thus, patients with diabetes have an expanded total body sodium pool, regardless of whether they are normotensive or hypertensive. Consequently, it does not appear to contribute directly to the mechanism of hypertension.

**Sympathetic activity**

Another system that contributes to the genesis of hypertension in individuals with diabetes is increased sympathetic nervous system activity. A direct correlation exists between increases in plasma insulin levels and sympathetic nervous system activity resulting from increased caloric intake [71]. Rowe and colleagues demonstrated insulin caused a dose-related increase in plasma norepinephrine levels whereas hyperglycemia had no such effects [72]. This increase in norepinephrine levels was closely related to an increase in pulse and blood pressure [72]. Thus, insulin may affect changes at the cellular receptor or synaptic level to mediate this increase in sympathetic activity. Consequently, hyperglycemia increases volume through sodium
retention and hyperinsulinemia increases sympathetic activity and together they generate sustained increases in blood pressure in people with poor glycemic control. Moreover, increases in sympathetic nervous system tone through drugs that result in vasoconstriction or sympathomimetics such as decongestants will worsen pre-existing insulin resistance or cause insulin resistance that is alleviated by renal denervation [73,74]. Vasoconstrictors such as norepinephrine, angiotensin II, vasopressin and many others are all growth factors whose action is modified by insulin [75]. The interaction of insulin with these various growth factors may relate to its vascular effects. Experimental studies have established that insulin can adrenergically mediate vascular smooth muscle contraction [76]. These effects of insulin are not related to its role in increasing glucose uptake because glucose removal from the incubation medium did not alter insulin’s effect. Conversely, the attenuating effect on vascular smooth muscle contraction was not observed in the presence of calcium channel antagonists and ouabain [77]. This is consistent with the concept that insulin decreases signal transduction by attenuating calcium influx across cell membranes. Moreover, insulin has been shown to directly alter both the action and cellular production of the vasoactive peptide, endothelin [78]. Unfortunately, in two separate clinical studies plasma concentrations of endothelin did not correlate with the presence of hypertension [79,80]. These observations further support the concept of cellular changes induced by insulin with no peripheral marker to reflect this action.

Insulin is well documented to induce different systemic and renal hemodynamic effects in euglycemic versus hyperglycemic individuals. Baron et al., examined the relationship between insulin and norepinephrine on blood pressure and insulin sensitivity in a group of lean and obese men [81]. Both groups were normotensive; however, the obese group had higher basal blood pressure values. They noted augmented pressor sensitivity and a decreased metabolic clearance to norepinephrine in obese compared to lean subjects. These studies suggest that abnormalities in beta adrenoreceptor activity are responsible for differences in hemodynamic effects of insulin. Moreover, epinephrine may potentially create a vicious cycle by inhibiting the effects of insulin and thus potentiating insulin resistance. Lastly, insulin resistance has been shown to alter pressor sensitivity in diabetic and obese nondiabetic individuals. However, it is not a direct cause of hypertension unless other genetic factors are also present.

Obesity is generally thought to be associated with increased sympathetic activity; however, an analysis between lean and obese hypertensives could not predict a difference in sympathetic activity [81–83]. Thus, this enhanced sympathetic activity in obese individuals may be simply due to overweight itself. In people with diabetes who have hypertension sympathetic activity is quite elevated compared to normotensive people with diabetes or those without diabetes [82,84]. Moreover, the level of sympathetic activity is proportional to the duration of diabetes [84]. Recent pilot studies of renal sympathetic nerve denervation demonstrated a clear improvement in blood pressure and glycemic control among patients with T2DM [85].

Therapeutic considerations

It is well established that in order to minimize cardio-renal risk of diabetes one has to maximally manage all known risk factors, that is, blood pressure, glucose, and lipids, Figure 74.1 [86].

Central to all CV risk factor modification are lifestyle changes, such as weight loss, exercise, reduction of alcohol, tobacco cessation, and a low-sodium diet below 2.4 g d⁻¹ (Table 74.1). Very few studies investigate the effects of sustained weight loss on hypertension. The Swedish Obese Subjects study followed patients for 8 years after an average weight loss of 20 kg following bariatric surgery. They found no significant improvement in blood pressure [87]. After smoking cessation, perhaps the most important is reduction of sodium and increasing potassium in the diet. Data clearly indicate that failure to reduce sodium intake will result in failure in most people to achieve blood pressure goals in spite of multiple medications [88]. This is due to increased sodium intake blunting the activity of the RAS and increasing sympathetic activity. Additionally, failure to achieve a serum potassium level of at least 3.8 mEq L⁻¹ will also blunt the antihypertensive activity of agents due to sustained vasoconstriction at low levels of potassium [89].

The goal blood pressure in patients with diabetes has recently come under scrutiny. For several years the American Diabetes Association and Joint National Committee (JNC VI and 7) have recommended a goal blood pressure of less than 130/80 mmHg in people with diabetes [90]. However, a review of all studies supporting this level of control reveals that all data come from retrospective analyses of trials and have a selection bias. Moreover, the only prospective outcome trials that randomized groups to different blood pressure levels were UKPDS and ACCORD but only the intensive treatment group in the ACCORD study attained a blood pressure of less than 130/80 mmHg.

In ACCORD, there was no significant difference in the primary endpoint, a composite of cardiovascular events, between the standard and intensive blood pressure groups, Figure 74.2 [91]. Moreover, there was a significantly higher side-effect profile in the intensive treatment group. Additional findings from post hoc analyses of diabetes subgroups of trials such as the Ongoing Telmisartan Alone and in combination with Ramipril Global Endpoint Trial (ONTARGET) and the International Verapamil SR-Trandolapril Study (INVEST) corroborate the blood pressure finding of ACCORD. Taken together, these three studies demonstrate no additional benefit of blood pressure lowering below 130/80 mmHg on CV risk reduction compared to being between 130–139/80–85 mmHg level [92,93]. INVEST also demonstrated an increase in CV events at systolic blood pressures <115 mmHg, although 100% of these patients had coronary artery disease [92].
Figure 74.1 Results of 13-year follow-up of intensive versus conventional risk factor follow-up in patients with T2DM on CV and renal outcomes. Source: Adapted from Baron 1994 [81].

The exception, however, is stroke reduction, which in ACCORD demonstrated a linear benefit between level of blood pressure and risk reduction, this was not seen in the INVEST analysis. Given that more recent guidelines will all be evidence based it is a reality that the blood pressure goal for those with diabetes will probably be <140/90 mmHg.

There are a number of classes of antihypertensive agents, and studies have shown some classes should be preferentially used in patients with diabetes. Many post hoc analyses demonstrate that diuretics and β-blockers both worsen blood glucose control among those with diabetes and increase the development of new-onset diabetes in those with impaired fasting glucose [94–96]. Thiazide diuretics worsen glycemic status through hypokalemia and other mechanisms related to increased visceral adiposity [97]. Vasoconstricting β-blockers worsen insulin sensitivity [98]. However, the vasodilating β-blockers, such as carvedilol and nebivolol have neutral effects on glycemic control and increase insulin sensitivity [99–101].

Although diuretics are known to worsen glycemic control, post hoc analyses of two different cardiovascular outcome trials noted that cardiovascular event rates were not higher [102,103]. An analysis of the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT) subgroup with diabetes failed to show a higher cardiovascular event rate in the diuretic group even though they had the greatest worsening of glycemic control [103]. In a more recent 12-year follow-up of the ALLHAT, participants on chlorthalidone with incident diabetes had consistently lower, nonsignificant risk for CVD mortality versus no diabetes or participants on amlodipine or lisinopril with incident diabetes. Moreover, participants with incident diabetes had elevated CHD risk compared with those with no diabetes, but those on chlorthalidone had significantly lower risk than those on lisinopril. Thus, thiazide-related incident diabetes has less adverse long-term CVD impact than incident diabetes that develops while on other antihypertensive medications [104].
Table 74.1  Lifestyle modifications to prevent and manage hypertension

<table>
<thead>
<tr>
<th>Modality</th>
<th>Modification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight reduction</td>
<td>Maintain normal body weight (body mass index 18.5–24.9 kg/m²).</td>
</tr>
<tr>
<td>Adopt DASH eating plan</td>
<td>Consume a diet rich in fruits, vegetables, and low-fat dairy products with a reduced content of saturated and total fat.</td>
</tr>
<tr>
<td>Dietary sodium reduction</td>
<td>Reduce dietary sodium intake to no more than 100 mmol per day (2.4 g sodium or 6 g sodium chloride).</td>
</tr>
<tr>
<td>Physical activity</td>
<td>Engage in regular aerobic physical activity such as brisk walking (at least 30 minutes per day, most days of the week).</td>
</tr>
<tr>
<td>Moderate alcohol use</td>
<td>Limit consumption to no more than 2 drinks (e.g., 24 oz beer, 10 oz wine, or 3 oz 80-proof whiskey) per day in most men and to no more than 1 drink per day in women and lighter weight persons.</td>
</tr>
</tbody>
</table>


Angiotensin-converting enzyme (ACE) inhibitors, angiotensin II receptor blockers (ARBs), and calcium channel blockers (CCBs) have beneficial or neutral effects on insulin sensitivity and glycemic control [94,98,105]. However, renin-angiotensin system (RAS) blockers administered concomitantly with thiazide diuretics do not prevent worsening of glycemic control in obese persons with impaired fasting glucose [106].

The presence of microalbuminuria (>30<300 mg d⁻¹) is associated with higher rates of cardiovascular disease [107–109]. Moreover, presence of macroalbuminuria (greater than 300 mg d⁻¹) in patients with diabetes is associated with a very high cardiovascular mortality rate as well as high risk of CKD progression [110]. It should be noted, however, that albuminuria reduction in response to blood pressure lowering therapy is not a surrogate for improved renal outcomes as has been seen in many recent studies [111–114]. Thus, these individuals must achieve blood pressure goals in order to reduce CV risk. For this reason, all patients with diabetes should be evaluated annually for albuminuria and any change in albuminuria. RAS blockers alone and in combination with CCBs will reduce albuminuria but for maximal reduction in albuminuria, a nondihydropyridine CCB (diltiazem, verapamil) should be used in concert with a RAS blocker to achieve BP goal [2,115].

To further support the concept that albuminuria reduction may not correlate with better renal outcomes, the results of the VA NEPHRON-D are instructive [116]. The combination of ACEI and ARBs were evaluated on renal outcomes in patients with diabetic nephropathy in this trial. This study investigated the combination of losartan and lisinopril in patients with T2DM with a GFR of 30–89 mL min⁻¹ per 1.73 m² and a urinary albumin-to-creatinine ratio of at least 300 mg g⁻¹ of Cr. The subjects treated with combination losartan and Lisinopril compared to losartan and placebo had higher rates of acute kidney injury and hyperkalemia. Also, the combination ARB/ACEI group did not result in a significant reduction in the primary endpoints of renal-disease progression or death [116].

Recent evidence from the international kidney guidelines (KDIGO) indicates that the strongest level of evidence supporting RAS blocker associated slowing of CKD progression in diabetic nephropathy is restricted to those with stage 3 or higher CKD who have albuminuria >300 mg day⁻¹ [117]. There is no evidence supporting their use in people with early kidney disease who are at risk of chronic kidney disease.

Figure 74.2  A suggested approach to achieve BP goal in patients with diabetes with Stages 1–3a chronic kidney disease. Source: Bakris 2010 [128]. Reproduced with permission of Elsevier.
nephropathy or those with microalbuminuria with or without hypertension [118]. Thus, the true evidence for blockers of the RAS is limited in people with diabetes. This is further supported by a meta-analysis indicating that BP lowering and not the class of antihypertensive agent used to reduce BP is responsible for a reduction in cardiovascular events [119]. Therefore, one should focus on lowering blood pressure using antihypertensive agents that do not worsen preexisting metabolic conditions and are shown to protect against kidney disease progression, that is, albuminuric kidney disease.

Combination therapy is commonplace in people with diabetes with or without kidney disease. An analysis of the diabetes subgroups of many cardiovascular outcome trials indicates an average of 2.9 medications are needed to achieve blood pressure goals. In subjects with stage 3 or greater kidney disease, this number rose to 3.3 medications [120]. Based on the American Society of Hypertension consensus report on the use of combination antihypertensive medications, a suggested approach to blood pressure control in patients with diabetes is detailed in Figure 74.2 [121]. This algorithm serves as a guide and does not substitute good clinical judgment for each unique patient.

Use of single pill combinations (SPC) also improves patient adherence and minimizes the number of BP pills ingested daily [122,123]. Fundamentally, two types of SPCs have good evidence for outcomes with one showing superior CV outcomes in diabetes, that is, RAS blockers with CCBs and RAS blockers with thiazide diuretics [121]. Between these two combinations RAS blockers with ACE inhibitors have level 2A [124]. Based on these data use of SPC should be strongly considered especially in people who are 20/10 mmHg above their BP goal [125]. Note that use of an ACE inhibitor/ARB combination is not supported for BP or CV risk reduction and is contraindicated in people with advanced diabetic nephropathy due to high rates of hyperkalemia [121].

**Conclusion**

It appears that development of hypertension in subjects with diabetes is due to both genetic and environmental modifiers of these genetic factors. The presence of insulin resistance should be considered a modifiable risk factor for, rather than the mechanism for development of hypertension in such patients. Insulin resistance can be markedly improved by weight reduction and exercise. Moreover, it is clear that an as yet undefined genotype needs to be present for the development of hypertension independent of that for diabetes. Thus, while obesity is clearly associated with insulin resistance and elevations in blood pressure, frank hypertension may not develop unless a certain “genetic load” is present. Major factors, but not the only contributors to hypertension development in this setting, include hyperglycemia-associated sodium retention/volume expansion and hyperinsulinemia-associated increases in sympathetic tone.

The ionic transporters primarily implicated in the genesis of hypertension involve calcium homeostasis. These transporters are affected by hyperglycemia as well as hyperinsulinemia. Cellular changes in these transporters ultimately affect sodium homeostasis. However, sodium retention, per se, is not a prominent factor that contributes to the genesis of hypertension in diabetes but increased salt sensitivity due to a reduction in nitric oxide is a factor. Thus, altered vascular responsiveness to sodium loads in such patients plays an important role in its genesis. Lastly, the vascular reactivity effects of sodium coupled with the enhanced pressor dose–response seen in subjects with diabetes combine to raise blood pressure in such individuals. These blood pressure effects early in the course of diabetes are largely vascular in nature with evidence of abnormal sodium handling by the kidney as noted earlier in this chapter. As eGFR falls below 60 mL min⁻¹ however, the kidney itself is further contributing to increasing blood pressure due to enhanced effects associated with increased sympathetic tone and sodium reabsorption [83,126,127].

The high cardiovascular risk in these patients requires an integrated therapeutic intervention that includes both lifestyle changes, appropriate medications for BP lowering, along with glucose and lipid control. The medications used to lower blood pressure should address albuminuria reduction, if present, and avoid worsening of glycemic control if possible. The most important aspect of blood pressure control in people with diabetes is actually lowering the blood pressure to goal, for example <140/90 mmHg while minimizing side effects.

**References**

Arterial hypertension in diabetes: etiology and treatment


Peripheral vascular and cerebrovascular disease in diabetes mellitus

Debabrata Mukherjee
Texas Tech University, El Paso, TX, USA

Key points

- Diabetes is frequently associated with atherosclerotic vascular disease including coronary, peripheral, and cerebrovascular disease.
- Early diagnosis of PAD in diabetic patients is critically important for the prevention of progression of disease as well as for prediction and subsequent reduction of overall cardiovascular risk.
- The American Diabetes Association (ADA) consensus statement recommends that a screening ABI be performed in all diabetic individuals >50 years of age.
- Current evidence highlights the importance of adopting a multifactorial approach for the prevention of vascular complications in patients with type 2 diabetes. Lowering of blood pressure with regimens based on a variety of antihypertensive drugs, including ACE inhibitors, ARBs, β-blockers, diuretics, and calcium channel blockers, has been shown to be effective in reducing cardiovascular events in diabetics.
- Cardiovascular physicians should be aware of the strong association between diabetes and atherosclerosis and use appropriate medical and interventional treatments to reduce disability and death in these patients.

Introduction

Diabetes mellitus is an established risk factor for atherosclerosis, and the risk of atherosclerotic vascular disease and its major clinical consequences, which include coronary artery disease (CAD), peripheral arterial disease (PAD), and cerebrovascular disease, is markedly increased among individuals with diabetes. The increased risk appears to be independent of, and additive to, other cardiovascular risk factors [1]. Epidemiologic studies suggest that atherosclerosis causes most of the morbidity and mortality in patients with diabetes, particularly in the burgeoning type 2 diabetic patient population [2]. The Verona Diabetes Study demonstrated that cardiovascular disease is responsible for 44% of all-cause mortality in the diabetic patient population [3]. The duration of diabetes increases the risk of death from cardiovascular disease, independent of coexisting risk factors. Insulin resistance and its attendant metabolic abnormalities appear to play a pivotal role in the pathophysiology of the increased cardiovascular risk of diabetes [4].

Diabetes and risk of vascular disease

A meta-analysis of individual records of diabetes, fasting blood glucose concentration, and other risk factors in people without initial vascular disease attempted to quantify the magnitude of associations of diabetes mellitus and fasting glucose concentration with risk of coronary heart disease and major stroke subtypes [5]. Analyses included data for 698,782 people (52,765 nonfatal or fatal vascular outcomes; 8.49 million person-years at risk) from 102 prospective studies. Adjusted HRs with diabetes were: 2.00 (95% CI 1.83–2.19) for coronary heart disease; 2.27 (1.95–2.65) for ischemic stroke; 1.56 (1.19–2.05) for hemorrhagic stroke; 1.84 (1.59–2.13) for unclassified stroke; and 1.73 (1.51–1.98) for the aggregate of other vascular deaths. Overall, it appears that diabetes confers about a twofold excess risk for a wide range of vascular diseases, independent from other conventional risk factors [5]. In T2DM, both angiogenesis and microangiopathy are increased and may contribute to accelerated atherosclerosis and the development of vulnerable plaque [5]. Hyperglycemia is a driving force in both large- and small-vessel disease [6].

Diabetes and peripheral arterial disease

Peripheral arterial disease (PAD) affects approximately 12 million people in the US and approximately 20–30% of these
patients have diabetes [1]. In studies using the ankle–brachial index (ABI), the prevalence of PAD (defined as an ABI <0.90) in diabetic individuals ranges from 20% to 30% [7]. Overall, diabetes is associated with a two- to fourfold increase in the incidence of PAD and an abnormal ABI is present in ~15% of diabetes patients [2,8]. Both intermittent claudication and critical limb ischemia are increased in diabetes. The prevalence of PAD increases with advancing age, duration of diabetes, and other risk factors such as smoking, dyslipidemia, and hypertension. The degree of diabetic control is also an independent risk factor for PAD; with every 1% increase in glycosylated hemoglobin, the risk of PAD has been shown to increase by 28% [9]. Early diagnosis of PAD in diabetic patients is critically important for the prevention of progression of disease as well as for prediction and subsequent reduction of overall cardiovascular risk. The distribution of atherosclerosis differs in diabetics and nondiabetics with stenotic lesions in patients with diabetes often located more distally than in nondiabetic subjects. Thus, the typical diabetic PAD lesions are located in the popliteal artery or in the runoff vessels below the knee, that is, the anterior tibial, posterior tibial, and the peroneal arteries [10,11]. The involvement of the distal limb vessels, such as the tibial and peroneal arteries, limits the potential for collateral vessel development and reduces options for revascularization [10].

Another hallmark of diabetic PAD is the calcification of the media layer of the arterial wall which may confound some diagnostic tests, a highly characteristic feature of T2DM [12,13]. A patient with critical limb ischemia is defined as a patient with chronic ischemic rest pain, ulcers, and gangrene due to arterial disease. It is important to consider that ulcers may often exist in the diabetic foot despite a normal macrocirculation. These ulcers may be due to disease in the microcirculation or related to neuropathy or sometimes multifactorial due to combination of impaired circulation, neuropathy, and infection.

### Diagnosis
A thorough medical history and physical examination is indicated in evaluating a diabetic individual for the presence of PAD. Information about the onset and duration of symptoms, pain characteristics, and any alleviating factors is helpful. Symptoms of leg ischemia in diabetic patients with peripheral neuropathy may be atypical and may delay diagnosis. Rather than experiencing cramping pain in legs or typical claudication, the patient may suffer from leg fatigue or inability to walk at their normal speed. If the patient has typical claudication, the claudication distance should be recorded at each visit and a shortening claudication distance signals progression of disease. The clinical stage of symptomatic PAD can be classified using the Fontaine staging system (Table 75.1). Fontaine stage I represents asymptomatic PAD; stages Ila and IIB include patients with mild and moderate-to-severe intermittent claudication, respectively; those with ischemic rest pain are classified as Fontaine stage III; and patients with tissue breakdown in the form of distal ulceration and gangrene represent Fontaine stage IV. A typical history of claudication has a low sensitivity, but a high specificity for PAD in diabetic individuals [14].

A complete physical examination is very important for the diagnosis of PAD in these individuals. Palpation of pulses in the leg and visual inspection of the feet are essential. Dependent rubor, pallor when the foot is elevated, absence of hair growth, and dystrophic toenails are signs of peripheral ischemia. In addition to measurement of ABI, physical examination should include blood pressure measurement, palpation of peripheral pulses, and auscultation of pulses and bruits. Palpation of peripheral pulses should include an assessment of the femoral, popliteal, and pedal vessels and pulses graded as absent, diminished, or normal. Dorsalis pedis pulse abnormalities are less sensitive for PAD, since up to 30% of these abnormalities may be due to a congenital absence of the dorsalis pedis artery. The absence of both the dorsalis pedis pulse and the posterior tibial pulse strongly suggests the presence of PAD. Figure 75.1 shows a typical protocol for the diagnosis of PAD in patients with diabetes.

An objective measure of peripheral vascular disease is the ABI, defined as the ratio between the arterial pressure at the ankle level (dorsalis pedis or posterior tibial) and in the left or right brachial artery with the highest pressure. The ABI should normally be above 0.9. This measurement is valuable for early detection of PAD and also for a better stratification of the overall cardiovascular risk. An ABI below 0.5 is indicative of severely impaired circulation of the foot. An ABI >1.4 is also abnormal and indicates poorly compressible vessels as a result of stiff arterial walls, which usually in diabetic patients are due to atherosclerosis in the media layer of the arterial wall. In situations where an elevated ABI is recorded or a pseudonormal value is suspected, the blood pressure should also be measured at the level of the toe by a minicuff and a technique suitable for blood flow detection in the toe. This is called the toe–brachial index (TBI).

The American Diabetes Association (ADA) consensus statement recommends that a screening ABI be performed in all diabetic individuals >50 years of age [1]. If normal (0.91–1.40), the test should be repeated every 5 years. An ABI should also be performed in any patient with symptoms suggestive of PAD. It should be recognized that ABI determinations may be of limited value in some patients with diabetes, because calcification of the tibial arteries may render them noncompressible, resulting in unusually high ABI values (>1.40). Under these conditions, the ABI cannot distinguish patients who have arterial occlusion from those who do not, making the ABI unhelpful. As a result,

---

**Table 75.1** Fontaine classification of peripheral arterial disease

<table>
<thead>
<tr>
<th>Stage</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Asymptomatic</td>
</tr>
<tr>
<td>Ila</td>
<td>Mild claudication (&gt;200 M)</td>
</tr>
<tr>
<td>IIB</td>
<td>Moderate to severe claudication (≤200 M)</td>
</tr>
<tr>
<td>III</td>
<td>Ischemic rest pain</td>
</tr>
<tr>
<td>IV</td>
<td>Tissue loss or ulceration</td>
</tr>
</tbody>
</table>
measurement of great toe artery pressure for calculation of TBI is commonly advocated in diabetic patients. Studies have suggested that assessment of TBI is the method of choice in the presence of overt calcification as defined by an ABI of >1.4 [15,16]. However, an elevated ABI is still predictive of an increased risk of cardiovascular events, and other noninvasive vascular tests should be considered to make the diagnosis of PAD. Transcutaneous oxygen (TcPO₂) measurement is another useful noninvasive modality that can prospectively determine severity of foot ischemia, and aid in selecting appropriate treatment for patients with diabetes and foot salvage problems [17].

In the patient with PAD in whom further investigation is required, for example in planning a revascularization procedure, the next step would be evaluation for segmental pressure and pulse volume recordings. Both tests aid in the localization of arterial occlusive lesions. Other noninvasive imaging techniques, such as ultrasonic duplex scanning or magnetic resonance angiography (MRA), can be used when more precise measurements of the morphological features of occlusions are required for planning revascularization options. Ultrasound duplex scanning can directly visualize vessels, providing information on artery wall thickness, degree of flow turbulence, and changes in blood flow velocity which can be used to define severity of arterial stenosis [18]. Comprehensive imaging of the peripheral vasculature has traditionally been performed with invasive digital subtraction angiography. However, with the introduction of MRA and computed tomographic angiography (CTA), noninvasive imaging is now becoming a reality. Contrast-enhanced MRA produces images that are fairly comparable with conventional angiography. Recently, the resolution of CTA has dramatically improved image quality and expanded the applications for noninvasive angiography. At this point, CTA is replacing conventional angiography in some centers. Table 75.2 depicts the different methods for evaluating the peripheral circulation.

**Figure 75.1** Typical protocol for the diagnosis of peripheral arterial disease in patients with diabetes. Source: Hiatt 2001 [14]. Reproduced with permission of Massachusetts Medical Society.

**Treatment**

One aim of medical management of PAD among diabetics is to aggressively modify cardiovascular risk factors to reduce the risk of future cardiovascular events (Table 75.3). It is also important...
to relieve the symptoms of intermittent claudication in order to improve functional status and quality of life. Smoking cessation, normalization of lipid abnormalities, and optimization of glycemic control are the most important steps that contribute to the prevention of primary or secondary cardiovascular disease in these patients. Table 75.3 describes general recommendations for treatment of diabetic patients with PAD.

PAD patients with diabetes tend to have more severe symptoms and a worse prognosis than nondiabetic patients [10], and the ACC/AHA guidelines concur with the American Diabetes Association (ADA) recommendations and note that aggressive management of diabetes has been shown to reduce the risk of nephropathy, retinopathy, and other microvascular events [19]. Intensive glycemic control may also reduce the risk of cardiovascular events in patients with PAD and diabetes, although there is a lack of clinical data specific to this population [19].

The Steno-2 study showed that the risk of both cardiovascular and microvascular complications was significantly reduced by about 50% in type 2 diabetic patients who received targeted, intensified, multifactorial management of glycemia, dyslipidemia, hypertension and microalbuminuria via lifestyle modification, pharmacologic therapy and secondary prevention with antiplatelet therapy, compared with patients conventionally managed according to national guideline recommendations [20]. Analysis of the Diabetes Control and Complications Trial, in which 1441 patients with T1DM were randomized to intensive or conventional treatment, showed that the risk of major peripheral vascular events in the group receiving intensive therapy was 22% lower than that in the conventional therapy group (0.43 vs. 0.55 events per 100 patient-years). This difference, however, was not statistically significant [21].

Glycemia

The UKPDS 33 study of 3867 newly diagnosed patients with T2DM showed that intensive antidiabetic treatment reduced the risk of complications by 12% (95% CI 1–21; p = 0.029) compared with conventional therapy [22]. Although this was primarily due to a reduction in microvascular complications, there was also a 16% reduction in MI (p = 0.052) with intensive treatment. The UKPDS long-term trials indicate that in patients with T2DM, intensive glucose control may have a legacy effect, whereas the same benefit may not apply to tight blood pressure control [23]. The studies highlight the importance of adopting a multifactorial approach for the prevention of vascular complications in patients with T2DM and reinforce the importance of maintaining good glycemic control, not only for the prevention of the renal and metabolic complications of diabetes but also for protection against the development of major cardiovascular disease in the long term [23]. A recent study, the Veterans Affairs Diabetes Trial (VADT) suggested that intensive glucose control in patients with poorly controlled T2DM had no significant effect on the rates of major cardiovascular events, death, or microvascular complications, with the exception of albuminuria [24]. The study randomly assigned 1791 military veterans (mean age, 60.4 years) who had a suboptimal response to therapy for T2DM to receive either intensive or standard glucose control. The median follow-up was 5.6 years. Median glycated hemoglobin levels were 8.4% in the standard-therapy group and 6.9% in the intensive-therapy group. The primary outcome occurred in 264 patients in the standard-therapy group and 235 patients in the intensive-therapy group (hazard ratio (HR) in the intensive-therapy group, 0.88; 95% CI 0.74–1.05; p = 0.14). There was no significant difference between the two groups in any component of the primary outcome or in the rate of death from any cause (HR 1.07; 95% CI 0.81–1.42; p = 0.62) [24]. The Action to Control Cardiovascular Risk in
Diabetes (ACCORD) trial, which involved 10,251 patients with T2DM, attempted to determine whether intensive insulin therapy was associated with a lower incidence of cardiovascular events than standard therapy [25]. The study was specifically designed to address whether an HbA1c goal of <6% to be attained by intensive therapy, would reduce cardiovascular events in patients with established cardiovascular disease or cardiovascular risk factors, as compared to a standard strategy using an HbA1c target of 7.0–7.9%. At a mean treatment duration of 3.5 years, the study was stopped prematurely on the recommendation of the Data and Safety Monitoring Board, owing to an increase in all-cause mortality in the intensive-therapy group compared with the standard-therapy group (5% vs. 4%; HR 1.22; 95% CI 1.01–1.46). The rate of death from cardiovascular causes was similarly increased in the intensive-therapy group (2.6% vs. 1.8%; HR 1.35; 95% CI 1.04–1.76). The primary outcome of nonfatal myocardial infarction, nonfatal stroke or death from cardiovascular cause was not significantly different in the group assigned to intensive glycemic control (6.9% vs. 7.2%; HR 0.90; 95% CI 0.78–1.04) [25]. The Action in Diabetes and Vascular Disease: Preterax and Diamicron Modified Release Controlled Evaluation (ADVANCE) trial enrolled 11,140 patients with T2DM and pre-existing cardiovascular disease or at least one additional cardiovascular risk factor [26]. In contrast to the ACCORD trial, the primary outcome in the ADVANCE trial was a composite of microvascular events (nephropathy and retinopathy) and macrovascular disease defined by major adverse cardiovascular events (myocardial infarction, stroke, and cardiovascular death). Intensive therapy was associated with a decrease in the incidence of the primary endpoint (a combination of microvascular and macrovascular events, as outlined earlier; 18.1% vs. 20.0%; HR 0.90; 95% CI 0.82–0.98) and of major microvascular events (9.4% vs. 10.9%; HR 0.86; 95% CI 0.77–0.97), compared with standard therapy. This benefit was primarily the result of a reduction in the incidence of nephropathy (4.1% vs. 5.2%; HR 0.79; 95% CI 0.66–0.93), as intensive therapy had no statistically significant effect on the incidence of retinopathy or macrovascular events [26].

Despite the need for further study on the association of glycemic control and cardiovascular risk in patients with PAD and diabetes, both the ACC/AHA and the ADA recommend the reduction of HbA1c to <7.0% for this population [19,27,28]. The results from ACCORD, ADVANCE, and VADT should not be interpreted to abandon the general goal of <7% and diminish the importance of glycemic control. The lower-than-anticipated event rates observed in both intensive and standard treatment groups in these studies rather support the premise that a multifactorial approach should be used to address the major cardiovascular risk factors including regular physical activity, lipid lowering, blood pressure control, and so on.

Antiplatelet therapy
The use of antiplatelet agents is known to reduce future secondary cardiovascular events in patients with both diabetes mellitus and cardiovascular disease [29]. One study examined whether aspirin and antioxidant therapy, combined or alone, are more effective than placebo in reducing the development of cardiovascular events in patients with diabetes mellitus and asymptomatic PAD. A total of 1276 adults aged 40 or more with T1 or T2DM and an ankle–brachial pressure index of 0.99 or less but no symptomatic cardiovascular disease were enrolled [30]. Daily, 100 mg aspirin tablet plus antioxidant capsule (n = 320), aspirin tablet plus placebo capsule (n = 318), placebo tablet plus antioxidant caplide (n = 320), or placebo tablet plus placebo capsule (n = 318). Overall, 116 of 638 primary events occurred in the aspirin groups compared with 117 of 638 in the no aspirin groups (18.2% vs. 18.3%) (HR 0.98; 95% CI 0.76–1.26) [30]. This trial did not provide evidence to support the use of aspirin in primary prevention of cardiovascular events and mortality in the population with diabetes. Based on this and other available evidence, the American Diabetes Association suggests that low-dose (75–162 mg d−1) aspirin use for prevention is reasonable for adults with diabetes and no previous history of vascular disease who are at increased risk for cardiovascular disease (10-year risk of CVD events >10%) and states that aspirin should not be recommended for CVD prevention for adults with diabetes at low CVD risk (men under age 50 years and women under 60 years with no major additional CVD risk factors [31]. The Antithrombotic Trialists’ (ATT) Collaboration performed a meta-analysis and suggested that in primary prevention without previous disease, aspirin is of uncertain net value as the reduction in occlusive events needs to be weighed against any increase in major bleeds [32].

Lipids
The Cholesterol Treatment Trials’ (CTT) Collaborators analyzed data from 18,686 individuals with diabetes (1466 with type 1 and 17,220 with type 2) in the context of a further 71,370 without diabetes in 14 randomized trials of statin therapy [33]. During a mean follow-up of 4.3 years, there were 3247 major vascular events in people with diabetes. There was a 9% proportional reduction in all-cause mortality per mmolL−1 reduction in LDL-cholesterol in participants with diabetes (rate ratio (RR) 0.91; 95% CI 0.82–1.01; p = 0.02), which was similar to the 13% reduction in those without diabetes (RR 0.87; 0.82–0.92; p < 0.0001). This finding reflected a significant reduction in vascular mortality (RR 0.87; 0.76–1.00; p = 0.008) and no effect on nonvascular mortality (RR 0.97; 0.82–1.16; p = 0.7) in participants with diabetes. There was a significant 21% proportional reduction in major vascular events per mmolL−1 reduction in LDL-cholesterol in people with diabetes (RR 0.79; 0.72–0.86; p < 0.0001), which was similar to the effect observed in those without diabetes (RR 0.79; 0.76–0.82; p < 0.0001). In diabetic participants there were reductions in myocardial infarction or coronary death (RR 0.78; 0.69–0.87; p < 0.0001), coronary revascularization (RR 0.75; 0.64–0.88; p < 0.0001), and stroke (RR 0.79; 0.67–0.93; p = 0.0002). These findings would suggest that statin therapy be considered for all
diabetic individuals who are at sufficiently high risk of vascular events [33]. However, it should be noted that while statins are effective for cardiovascular disease prevention, they have recently been associated with an increased risk of new-onset diabetes mellitus [34,35]. Based on new data, the FDA changed statin labeling, incorporating the fact that there are studies showing that patients being treated with statins may have a small increased risk of increased blood sugar levels and of being diagnosed with T2DM. Based on totality of data, for the vast majority of patients who are on statins, the benefits are expected to outweigh the risks since statins are very effective at lowering risk for vascular disease and stroke. The Justification for the Use of Statins in Prevention: an Intervention Trial Evaluating Rosuvastatin (JUPITER) trial investigated whether treatment with rosuvastatin, 20 mg daily, as compared with placebo, would decrease the rate of first major cardiovascular events in apparently healthy men and women with LDL-cholesterol levels of less than 130 mg dL\(^{-1}\) (3.4 mmol L\(^{-1}\)) and high-sensitivity C-reactive protein levels of 2.0 mg L\(^{-1}\) or higher [36]. The study reported that rosuvastatin significantly reduced the incidence of major cardiovascular events [36]. The concern about new-onset diabetes should lead to open discussion between physician and patient and the risk–benefit assessed for each individual patient. It should also lead to increased vigilance about testing for diabetes in patients who are on statins.

Fibrates, introduced more than 35 years ago, based on favorable changes in the lipid profile, continue to generate controversy regarding their clinical efficacy. Two randomized, placebo-controlled trials of gemfibrozil had demonstrated improvements in cardiovascular outcomes but subsequent trials of bezafibrate and fenofibrate showed no significant overall cardiovascular benefit over placebo [37]. Based on contemporary evidence, it would appear that the benefit of adding a fibrate to statin therapy in reducing the risk of cardiovascular events in patients with T2DM is unproven. At this time, clinicians who choose to prescribe combination therapy should selectively target high-risk patients only after optimal control of LDL-cholesterol has been achieved with statin therapy [37].

**Blood pressure**

Hypertension is a common comorbidity of diabetes, affecting a significant proportion of patients, with prevalence depending on type of diabetes, age, obesity, and ethnicity. In T1DM, hypertension is often the result of underlying nephropathy, while in T2DM it usually coexists with other cardiometabolic risk factors. Epidemiologic analyses show that high blood pressure is associated with increased cardiovascular event rates and mortality in individuals with diabetes. Randomized clinical trials have demonstrated the benefit (reduction of CHD events, stroke, and nephropathy) of lowering blood pressure to <140 mmHg systolic and <80 mmHg diastolic in individuals with diabetes but the evidence for benefits from lower systolic blood pressure targets is limited [38]. The current guidelines recommend that people with diabetes and hypertension should be treated to a systolic blood pressure goal of <140 mmHg and diastolic blood pressure <80 mmHg [38]. Lowering of blood pressure with regimens based on a variety of antihypertensive drugs, including ACE inhibitors, ARBs, β-blockers, diuretics, and calcium channel blockers, has been shown to be effective in reducing cardiovascular events in diabetics. Although evidence for distinct advantages of RAS inhibitors on cardiovascular disease outcomes in diabetes remains conflicting, the high CVD risks associated with diabetes, and the high prevalence of undiagnosed CVD, may favor using them as first-line hypertension therapy in people with diabetes [38].

**Lifestyle**

To relieve the symptoms of intermittent claudication, patients should exercise regularly. A regular walking regimen is extremely helpful. The best program is a stop-start walking regimen and includes regular daily walks, 30–45 min d\(^{-1}\), at least three times per week, for at least 6 months. Individuals should walk as far as possible using near maximal pain as a signal to stop and resume walking when pain goes away. A typical supervised exercise program is 60 min in duration and is monitored by a skilled nurse or technician. Patients may be encouraged to walk primarily on a treadmill since this most closely reproduces walking in the community setting. The initial workload of the treadmill is set to a speed and grade that bring on claudication pain within 3–5 min. Patients walk at this work rate until they achieve claudication of moderate severity. They then rest until the claudication abates, and then resume exercise. This repeated on-and-off form of exercise is continued throughout the supervised rehabilitation setting. On a weekly basis, patients should be reassessed clinically as they are able to walk further and further at their chosen workload. This then will necessitate an increase in speed or grade or both to allow patients to successfully work at harder and harder workloads [39–41]. Currently, two pharmacologic agents are approved for the symptomatic treatment of intermittent claudication: namely pentoxifylline and cilostazol. Pentoxifylline, a hemorheologic agent, decreases blood viscosity and improves erythrocyte flexibility [42]. The results of clinical trials demonstrating the efficacy of pentoxifylline in improving treadmill-walking distance have been equivocal, and there are insufficient data to justify generalized use in PAD [43]. Cilostazol, a phosphodiesterase inhibitor, is the most effective agent available in the US. Cilostazol (100 mg twice daily) has been shown to improve maximal walking distance by 40–50% compared with placebo [44,45]. In a direct comparison, the mean maximal walking distance in PAD patients treated with cilostazol for 24 weeks was significantly greater compared with that of patients who received pentoxifylline or placebo [46]. Because of concerns about the potential risk of mortality, cilostazol is contraindicated if any degree of systolic or diastolic heart failure is present or in patients with left ventricular systolic dysfunction [40].
Revascularization
Diabetic patients with progressively disabling claudication and those with critical limb ischemia should be considered for revascularization. Decisions about endovascular or open surgical procedures depend in large part on the severity and distribution of the arterial lesions [47]. Outcomes of iliac artery percutaneous transluminal angioplasty (PTA) and stenting in patients with diabetes have been reported as similar to or worse than those in nondiabetic patients [48,49]. With respect to surgery, the long-term patency rates after femoral-popliteal PTA are also lower in diabetic than in nondiabetic patients [50]. The modality of revascularization among diabetic patients with PAD should be tailored to the clinical circumstance, lesion characteristics, and patient preference. In the future, dedicated trials directly comparing endovascular versus surgical revascularization in diabetic patients with PAD may help define optimal treatment. It is important that patients undergoing either percutaneous or surgical revascularization receive optimal secondary preventative therapy post procedure with antiplatelet and lipid-lowering agents.

Diabetes and cerebrovascular disease
The risk of stroke and transient ischemic attacks (TIA) is significantly increased in patients with diabetes [51–54]. In fact, cerebrovascular disease is the most common long-term cause of morbidity and mortality in patients with both T1 and T2DM. Since initial observations by the Framingham investigators, several large population-based studies have confirmed an increased frequency of stroke in the diabetic population [54,55]. Diabetes was the strongest single risk factor for stroke (relative risk for men 3.4 and for women 4.9) in a prospective study from Finland with a follow-up of 15 years [56]. Among stroke subtypes, diabetes is a prominent risk factor for ischemic stroke, but data on hemorrhagic stroke have been conflicting. One study suggested that the risk of stroke among patients taking hypoglycemic medications was increased threefold among the nearly 350,000 men in the Multiple Risk Factor Intervention Trial [57]. In the Baltimore-Washington Cooperative Young Stroke Study, stroke risk increased more than 10-fold in diabetic patients younger than 44 years of age, ranging as high as 23-fold in young White men [58]. Diabetes also increases stroke-related mortality, doubles the rate of recurrent stroke, and triples the frequency of stroke-related dementia [59,60].

Diabetes may also cause microatheromas in small vessels, such as the lenticulostriate arteries, leading to lacunar stroke, a common subtype of ischemic stroke. Lacunar stroke is a unique subtype and requires specific clinical and imaging features for diagnosis. Stroke patients with diabetes, or with hyperglycemia in the acute stage of stroke, have a higher mortality, worse neurologic outcome, and more severe disability than those without [61]. There is less information concerning the risk of stroke in T1 than in T2DM. The World Health Organization Multinational Study of Vascular Disease in Diabetes reported increased cerebrovascular mortality in type 1 diabetic patients however, with considerable variations between countries [62]. The data from the nationwide cohort of more than 5000 Finnish childhood-onset type 1 diabetic patients showed that, by the age of 50 years, the risk for an acute stroke was equal to that of an acute coronary event without any gender-related differences [63]. Presence of diabetic nephropathy was the strongest predictor of stroke, causing a 10-fold increase of risk. After correction for other risk factors for stroke, which are also more common in diabetic subjects, the risk still remains increased more than twofold meaning that diabetes itself is a strong independent risk factor for stroke [64].

Prevention of stroke
Measures to prevent stroke in diabetes should include a multipronged strategy targeted at treatment of hypertension, hyperlipidemia, microalbuminuria, hyperglycemia, smoking cessation, and the appropriate use of antiplatelet medication (Table 75.4).

Blood pressure
Results from the HOPE Study and Perindopril Protection Against Recurrent Stroke Study (PROGRESS) suggest that the reduction of stroke incidence in diabetic subjects during treatment based on ACE-inhibitors was greater than would be anticipated from the blood pressure-lowering effect alone and the effect was also evident in normotensive individuals [65,66]. In the Losartan Intervention For Endpoint reduction in hypertension (LIFE) study the same trend was found with an angiotensin receptor blocker, losartan [67]. However, in several other trials, including the Anti-hypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT), there was no apparent benefit of one class of antihypertensive drug over another in this respect [68]. Current data suggests that optimal blood pressure lowering may be more important than a particular agent.

Lipids
Treatment with statins has been shown to reduce the incidence of stroke in high-risk patients, but the diabetic subpopulations in the trials have been too small to allow a reliable subgroup analysis. In the Heart Protection Study, a sizeable subgroup of 5963 diabetic patients was randomized to placebo or 40 mg of simvastatin daily. Simvastatin reduced the incidence of stroke by a robust 24% [69].

Antiplatelet therapy
Antiplatelet therapy has also been shown to reduce the incidence of stroke in diabetic patients and is indicated for both primary and secondary prevention of stroke [70]. Aspirin in a
Table 75.4 Recommendations for prevention and treatment of stroke in diabetic patients

<table>
<thead>
<tr>
<th>Recommendation</th>
<th>Classa</th>
<th>Levelb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normalization of blood pressure is recommended in all patients with diabetes</td>
<td>I</td>
<td>A</td>
</tr>
<tr>
<td>for the prevention of stroke</td>
<td></td>
<td></td>
</tr>
<tr>
<td>For stroke prevention, blood pressure lowering is more important than the</td>
<td>IIa</td>
<td>B</td>
</tr>
<tr>
<td>choice of drug</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inhibition of the renin-angiotensin-aldosterone system may have additional</td>
<td>IIa</td>
<td>B</td>
</tr>
<tr>
<td>benefits beyond blood pressure lowering per se</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inhibition of the renin-angiotensin-aldosterone system may be considered also</td>
<td>IIa</td>
<td>B</td>
</tr>
<tr>
<td>in diabetic patients with normal blood pressure levels</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patients with stroke should be treated with statins according to the same</td>
<td>I</td>
<td>B</td>
</tr>
<tr>
<td>principles as nondiabetic subjects with stroke</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antiplatelet therapy with aspirin is recommended for primary and secondary</td>
<td>I</td>
<td>B</td>
</tr>
<tr>
<td>prevention of stroke</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patients with acute stroke and diabetes should be treated according to the</td>
<td>IIa</td>
<td>C</td>
</tr>
<tr>
<td>same principles as stroke patients without diabetes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimization of metabolic conditions including glycemic control should be</td>
<td>IIa</td>
<td>C</td>
</tr>
<tr>
<td>aimed for</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

aClass of recommendation.  
bLevel of evidence.  
Source: Adapted from Ryden 2007 [80]. Reproduced with permission of Oxford University Press.

low dose (75–325 mg daily) should be the drug of initial choice. In patients with recurrent stroke, a combination of aspirin and dipyridamole may be considered [71]. The alternative combination with aspirin and clopidogrel seems less safe since it was associated with an increased risk of bleeding without any benefit in terms of cardiovascular outcome in the Management of ATherothrombosis with Clopidogrel in High-risk patients with recent TIA or ischemic stroke (MATCH) Trial, performed in 7599 patients of whom 68% had diabetes [72]. In patients with atrial fibrillation, anticoagulant therapy should be given for stroke prevention.

Revascularization

The high frequency of early stroke following TIA mandates a workup within 7 days of the index event to reduce the risk of a subsequent, and potentially more serious, and even fatal neurologic event. Initial evaluation with echocardiography and carotid ultrasound is usually indicated. After a TIA or stroke caused by carotid-artery disease, medical treatments can be optimized in high-risk patients, avoiding the need for emergency carotid surgery thus allowing patients to undergo safer elective surgery [73]. Carotid endarterectomy for the prevention of stroke in patients with high-grade stenosis of the carotid artery has been shown to be effective, although it has not been specifically investigated in diabetic patients. Since complications during and after this procedure are more frequent in diabetic as compared with nondiabetic subjects, special consideration should be given to the overall risk for peri- and postoperative morbidity and mortality when deciding on surgical interventions in the patient with diabetes [74]. The presence of diabetes, however, does not seem to increase the perioperative risk of stroke [75,76]. An alternative to endarterectomy, carotid artery angioplasty and stenting (CAS), which has been found to be at least not inferior to endarterectomy, may prove to be a preferable method in high-risk patients [77] but the effects of diabetes on carotid stenting have not been well studied.

Treatment of acute stroke

The treatment in the acute phase of stroke in diabetic patients should follow the same principles that govern the treatment of stroke in the general population. Available studies have not suggested any interaction of diabetes with treatment [78]. Thrombolysis is an effective treatment for ischemic stroke if instituted within 3 hours of symptom onset. It reduces mortality and disability from stroke but is associated with a risk of hemorrhage and its use and effects in diabetes require further evaluation in clinical studies. Conservative treatment of stroke includes close surveillance of vital signs, optimization of circulatory and metabolic conditions, including glycemic control, in a designated stroke unit. Patients should receive early neurologic rehabilitation with physical and occupational therapy to improve quality of life. Recent studies suggest that early intervention for hypertension during the acute phase of stroke may be beneficial but currently it is recommended to acutely reduce only very high blood pressures, above 220 mmHg systolic and/or 120 mmHg diastolic, and not lower blood pressure to levels that may enhance ischemia. Blood pressure should also not be lowered by more than 25% during the first day of treatment [79].
Conclusions

Noncoronary atherosclerotic disease such as PAD and cerebrovascular disease is a common finding in patients with diabetes. The risk of developing PAD and cerebrovascular disease is not only higher in patients with diabetes, but the disease is more severe and progresses aggressively than in nondiabetic individuals. In fact, diabetes is the most common cause of nontraumatic amputations in the United States. The major concern to patients with diabetes and PAD or cerebrovascular disease is from cardiovascular events, and the primary therapeutic goal is to modify and optimally treat atherosclerotic risk factors. Risk factor management in these individuals includes lifestyle modifications, treating associated conditions such as dyslipidemia and hypertension, and preventing ischemic events with antplatelet therapy [81]. Pharmacologic therapies to improve symptomatic PAD include cilostazol and pentoxiphylline. A supervised exercise program should be the initial treatment step for the management of symptomatic PAD prior to starting pharmacologic therapy. Revascularization has an important role to play in the management of patients with both PAD and cerebrovascular disease for whom risk factor modification and pharmacologic treatment prove inadequate. Cardiovascular physicians should be aware of the strong association between diabetes and atherosclerosis and use appropriate medical and intervention treatments to reduce disability and death in these patients. Dedicated prospective studies are indicated to define optimal antplatelet therapy and revascularization modality in diabetic patients.

References

22. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in...


35 Johansen KL: Increased diabetes mellitus risk with statin use: comment on “statin use and risk of diabetes mellitus in postmenopausal women in the Women's Health Initiative”. *Archives of Internal Medicine* 2012;172(2):152.


66 Randomised trial of a perindopril-based blood-pressure-lowering regimen among 6,105 individuals with previous stroke or transient ischaemic attack. Lancet 2001;358(9287):1033–1041.


Diabetes and public health
CHAPTER 76

The diabetes challenge: from human and social rights to the empowerment of people with diabetes

C.T. Di Iorio¹, Fabrizio Carinci¹ and M. Massi Benedetti²
¹Secretrix snc, Pescara, Italy
²Hub for International Health ReSearch HIRS, Italy

Key points

- A person with diabetes can experience a normal life.
- Discrimination still exists against people with diabetes and is widely diffused throughout the world.
- The right to enjoy the highest attainable standard of physical and mental health has still to be secured for people with diabetes.
- Strong action has been undertaken at international and local levels to protect the human and social rights of people with diabetes.
- The most relevant factor for the protection of the rights of the people with diabetes is their possibility to be an active member of their management team (empowerment).
- Empowerment can only be effective if based on solid and timely information.
- The right to accurate information is a prerequisite for the protection of human and social rights of people with diabetes.
- New technologies and methodologies have proven to be able to guarantee appropriate information in full and absolute respect of privacy.

Introduction

Diabetes is a growing "global epidemic" for which the major public health measures, including primary and secondary prevention, have basically failed to deliver the desired reduction of adverse outcomes that was globally expected.

Over 346 million people worldwide have diabetes, with a projected increase of 16% by 2025 [1]. According to the World Health Organization (WHO), 3.4 million people died in 2004 because of the consequences of high blood sugar. Unfortunately, such a threatening figure is expected to double by 2030 [2]. In economic terms, diabetes care may account for up to 15% of the national healthcare expenditure [3].

Diabetes has been defined as a “chronic, debilitating and costly disease associated with major complications that pose severe risks for families, countries and the entire world” [2]. The 2006 UN Resolution 61/225 [2] asks for “all nations to develop national policies for the prevention, care and treatment of diabetes … taking into account internationally agreed development goals, including the Millennium Development Goals.” Delegates participating at the WHO 65th World Health Assembly, “approved the development of a global monitoring framework for the prevention and control of non communicable diseases (NCDs), including indicators and a set of global targets.” At the same meeting, the Assembly approved a global target of a 25% reduction of deaths from NCDs by 2025.

Diabetes can play an important role in achieving such targets through an effort that will require substantial societal involvement. The diabetes challenge of the 21st century implies a shift from the “multifactorial disease approach” to a “multidimensional systemic vision” that will look simultaneously at all interrelated components. To achieve the WHO target will require a careful identification of global/national policies which must take into account all relevant environmental factors associated with the spread of the diabetes epidemic, including climate change, industry standards and consumer behavior, the health system organization and how it ensures an efficient collaboration through strategies of education and self-care.

At the same time, achieving targets in diabetes will need the active participation of people with diabetes.

The notion of empowerment is a relatively new element of a multifaceted strategy that has been recently embraced even at the highest political level [4]. It advocates the direct involvement of individuals in shared processes, by respecting human, social and health rights, with a continuous provision of accurate information and educational tools for disease management.

© 2015 John Wiley & Sons, Ltd. Published 2015 by John Wiley & Sons, Ltd.
According to the International Diabetes Federation (IDF), people with diabetes should live their lives as normally as possible. Only through the elimination of current barriers, they will be able to play an active and equal role in society, aware of the risk of major complications that could lead to disability, reduced quality of life, and death [5]. To guarantee the right to health to people living with diabetes, adequate consideration should also be given to all disabilities directly associated with the progression of the disease.

In this chapter, we provide an overview of the legal background and international instruments available to empower people with diabetes from a comprehensive perspective.

First, we present the fundamentals of the right to health through an overview of international treaties. Since 1948, many legislative instruments have been enacted at a global level, particularly by the United Nations (UN) and the European Union (EU).

The following section introduces the human and social rights of people with diabetes and provides details of the IDF International Charter of Rights and Responsibilities, which solemnly declares the current trend to enhance the role and responsibilities of individual subjects in the global fight against the disease.

The section “Rights and responsibilities on diabetes information and education” analyzes how such roles can be played in the context of the individual right to education and information, as required for the proper management of diabetes. The discussion herein aims to demonstrate that the adoption of a comprehensive notion of empowerment is currently recognized by the international instruments.

Finally, we envisage ways to implement such a comprehensive concept of empowerment. Strengthened healthcare systems require a deep understanding of the links between actions and different environmental factors. The case of interconnected diabetes registers is presented as a practical solution that could realize this vision and help meeting the ambitious targets recently set by the WHO.

The right to health

The right to health is enshrined in numerous international and regional human rights treaties. The 1948 UN Universal Declaration of Human Rights (Art. 25) first recognized that every human being has “the right to a standard of living adequate for the health and well-being of himself and of his family, including food, clothing, housing and medical care and necessary social services” [6]. However, the Declaration had no binding force until 1976, when the International Covenant on Civil and Political Rights (ICCPR) and the International Covenant on Economic, Social and Cultural Rights (ICESCR) became international law. The Universal Declaration of Human Rights, together with these two covenants, constitutes the so-called “International Bill of Human Rights.” Article 12 of the International Covenant on Economic, Social and Cultural Rights (1966) [7] specifically “recognize[s] the right of everyone to the enjoyment of the highest attainable standard of physical and mental health,” which includes an obligation on States Parties to take the necessary steps to fully realize the right to health.

The 2000 UN General Comment No.14 “Right to Health” [8] specifies that although Article 12 of the Covenant did not adopt the definition of health contained in the preamble to the Constitution of WHO, which envisages health as “a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity,” the reference in Article 12.1 of the Covenant to “the highest attainable standard of physical and mental health” has to be broadly interpreted. It means that the right to health should not be limited to the right to healthcare, but should extend to the underlying socioeconomic determinants of health, for example access to safe and potable water, adequate sanitation, adequate supply of safe food, nutrition, housing, healthy occupational and environmental conditions, access to health-related education and information on a range of topics, for example sexual and reproductive health. According to the General Comment, the right to health is composed of four elements: availability, accessibility, acceptability, and quality. Availability refers to the provision of sufficient public health and healthcare facilities, programs, goods and services. Accessibility is composed of four dimensions: nondiscrimination, physical accessibility, economical accessibility (affordability), and information accessibility. Acceptability implies that health facilities, goods and services must respect ethical principles. Quality means that health facilities, goods and services must be scientifically and medically sound.

The right to health has been also recognized in the following treaties: the 1965 International Convention on the Elimination of All Forms of Racial Discrimination [9], the 1979 Convention on the Elimination of All Forms of Discrimination against Women [10], and the 1989 Convention on the Rights of the Child [11]. At regional level, several human rights instruments also recognize the right to health.

In the EU, two international legal instruments play a crucial role in the recognition and affirmation of the right to health: the European Social Charter and the EU Charter of Fundamental Rights. The 1996 revised European Social Charter [12], which operates through international law, is binding for all signatory States, including all EU Member States (MS). Article 11 of the Charter proclaims the right to the protection of health and poses on Contracting Parties an obligation to take the necessary steps “to remove as far as possible the causes of ill-health; to provide advisory and educational facilities for the promotion of health and the encouragement of individual responsibility in matters of health; and to prevent as far as possible epidemic, endemic and other diseases as well as accidents.” The European Committee of Social Rights specified that the right to health includes...
the following components: ensuring an adequate healthcare system, access to health and healthcare for vulnerable groups, public health protection measures (e.g., preventing air and water pollution, noise abatement, food control, and environmental hygiene) and providing health education [13].

The EU Charter of Fundamental Rights [14] can be considered a major development in the EU agenda for human rights. The Charter has been incorporated in the Lisbon Treaty, which entered into force in 2009. As a consequence of the entry into force of the Lisbon Treaty, the Charter gained the same legal value as the EU Treaties. The catalogue of civil, political, economic and social rights enshrined in the Charter is currently legally binding for all MS, with regard to the implementation of Union law, and enforceable through the Court of Justice. The Charter expands the scope and recognition of fundamental rights, for example dignity, freedom, equality, solidarity, citizenship and justice, to the proclamation of additional rights such as data protection, bioethics, and the right to good administration. Article 35 of the Charter specifically deals with healthcare: “Everyone has the right of access to preventive health care and the right to benefit from medical treatment under the conditions established by national laws and practices. A high level of human health protection shall be ensured in the definition and implementation of all Union policies and activities.”

In the African Region, the 1981 African Charter on Human and Peoples’ Rights [15], which entered into force in 1986, recognizes the right to health in Article 16, which states that “every individual shall have the right to enjoy the best attainable state of physical and mental health. States parties to the present Charter shall take the necessary measures to protect the health of their people and to ensure that they receive medical attention when they are sick.”

Article 10 of the 1988 Additional Protocol to the American Convention on Human Rights [16] in the Area of Economic, Social and Cultural Rights (the Protocol of San Salvador) adopts the wording of the WHO Constitution, defining the right to health as “the enjoyment of the highest level of physical, mental and social well-being,” which has to be considered as a “social good.” Among measures ensuring the right to health, States Parties shall guarantee: primary healthcare; extension of the benefits of health services to all individuals subject to the State’s jurisdiction; universal immunization against the principal infectious diseases; prevention and treatment of endemic, occupational and other diseases; education of the population on the prevention and treatment of health problems, and satisfaction of the health needs of the highest risk groups and of those whose poverty makes them the most vulnerable.

Considering the above-mentioned international instruments it can be highlighted that the notion of right to health is globally conceived as inclusive of all the socioeconomic and environmental determinants of health and expands to various dimensions that are strictly dependent upon the realization of other fundamental human and social rights. As a matter of fact, the right to food, housing, work, education, human dignity, life, nondiscrimination, equality, the prohibition against torture, privacy, access to information, freedom of association, assembly and movement are strictly interlinked with the right to health and play a crucial role for achieving the highest attainable standard of health for all.

**Human and social rights of people with diabetes**

Diabetes is a chronic potentially debilitating disease leading to major complications as the disease progresses. Targeted measures and policies are needed to ensure that people with diabetes may enjoy their lives even in the face of significant hurdles.

The interdependence between the right to health and the other human and social rights is evident in the unfortunate occurrence of disability, where the respect of the principle of nondiscrimination is key to ensuring the highest attainable standard of health.

Many regional legislative instruments deal with persons with disabilities. The revised European Social Charter [12] states that persons with disabilities have the right to independence, social integration and participation in the life of the community (Art. 15); and States Parties have to undertake specific steps to ensure the effective exercise of this right. Article 26 of the EU Charter of Fundamental Rights [14] disciplines the integration of persons with disabilities, recognizing that they have to “benefit from measures designed to ensure their independence, social and occupational integration and participation in the life of the community.” The 1981 African Charter on Human and Peoples’ Rights (Art. 18) [15] specifies that “the aged and the disabled shall also have the right to special measures of protection in keeping with their physical or moral needs.” Article 10 of the 1988 Additional Protocol to the American Convention on Human Rights in the Area of Economic, Social and Cultural Rights (the Protocol of San Salvador) [16] provides “protection of the handicapped” by asserting that “everyone affected by a diminution of his physical or mental capacities is entitled to receive special attention designed to help him achieve the greatest possible development of his personality. The States Parties agree to adopt such measures as may be necessary for this purpose.”

Building upon previous legislative instruments, the UN General Assembly adopted the Convention on the Rights of Persons with Disabilities [17] in 2006. The Convention on the Rights of Persons with Disabilities and its Optional Protocol can be considered as a human rights instrument that provides explicit social development goals.

The Convention adopts a broad and evolving concept of disability. In particular, disability “results from the interaction between persons with impairments and attitudinal and environmental barriers that hinders their full and effective participation in society on an equal basis with others.”
It defines in detail the rights of people with disabilities and provides specific ways of implementation. Accordingly, States Parties have to guarantee a series of specific rights to persons with disabilities, spanning from the right to life on an equal basis with others, to the right to nondiscrimination, legal and social protection, privacy, education, work and health, including comprehensive habilitation and rehabilitation services in the areas of health, employment and education. Article 25 of the Convention deals with disabilities in relation to the attainment of the right to health, specifying that "persons with disabilities have the right to the enjoyment of the highest attainable standard of health without discrimination on the basis of disability."

According to Article 4, ratifying countries have to develop and implement policies, laws and administrative measures that ensure the fulfillment of the rights recognized by the Convention and repel incompatible laws/regulations. Among possible approaches to implementation, Article 8 also promotes a change in perceptions of disabled people, by increasing awareness of their capabilities and combating stereotypes and prejudices.

Finally, Articles 33–39 provide monitoring mechanisms to regularly assess progress in the implementation of the Convention.

Building upon fundamental human rights instruments, for example the Convention on the Rights of the Child (CRC), the Convention on the Elimination of All Forms of Discrimination against Women (CEDAW) and, above all, the Convention on the Rights of Persons with Disabilities, the IDF has developed the International Charter of Rights and Responsibilities of People with Diabetes [18]. The vision of the Charter is to "optimize health and quality of life, enable people with diabetes to have as normal a life as possible, reduce or eliminate the barriers which deny realization of full potential as members of society."

The Charter specifies the fundamental human and social rights of people living with diabetes, providing specific examples regarding the right to healthcare, information, education, and social justice, while at the same time recognizing their own responsibilities. Therefore, the Charter aims to guarantee the rights and freedoms of people living with diabetes, to combat the discrimination and stigma that they often face under such condition, and, at the same time, to foster their empowerment.

In particular, the right to care includes the following: to have an early diagnosis and affordable and equitable access to care and treatment; to receive regular, reliable advice, education and treatment in accordance with evidence-based practices; to benefit from proactive health sector community outreach, education and prevention campaigns in every healthcare setting; to access high-quality services and care during and after pregnancy and childbirth, during childhood and adolescence; to keep information related to their disease confidential and in accordance with privacy and data protection principles and to receive continuous and appropriate care in cases of disaster and emergency.

The right to information and education spans from the prevention, detection, management and access to educational resources, to the training in self-management, the direct involvement (together with healthcare organizations) in the planning of health goals, obtaining information on interventions (including risks and benefits), and access to medical records.

The right to social justice implies that people with diabetes shall be considered as fully engaged members of the society, who can access medication and monitoring technologies, are not discriminated against in accessing insurances and a driving license, are fairly treated in employment and career progression, and are duly supported in social and school/work environments.

Finally, people with diabetes have the responsibility to share all relevant information with healthcare providers, be able to manage their agreed care and treatment plan, implement healthy lifestyle behaviors as part of their self-management, inform family, school/work/social colleagues of their condition and be respectful of the rights of others with diabetes.

The Charter finally urges the international community to apply the principles therein contained, through national plans for diabetes care, prevention, research and education, while ensuring a regular monitoring and review of the Charter's status of implementation.

Although the Charter is not legally binding for the international community, nevertheless it shall be regarded as an international declaration that can be used as a powerful campaign instrument to help counteract the diabetes epidemic through a global concerted action.

**Rights and responsibilities on diabetes information and education**

As early as 1989, the St. Vincent Declaration [19] recognized that reducing the burden of diabetes could only be achieved through an active partnership between people with diabetes and service providers, particularly in the areas of disease management and education.

This assumption advocates the fundamental principles of patient empowerment, a concept that after two decades has become broadly used in the field of healthcare. Empowerment has been defined as “the ability of a person affected by a disease to be an active member of his/her management team” [20]. It addresses different dimensions of care management, for example the ability of a person to make decisions on treatments, education on both medical and health conditions, as well as consciousness of the emotional impact of the disease. Patient empowerment integrates multiple concepts allowing a person to effectively self-manage her/his disease [21]. To make this possible, patients have to be involved and educated on the application of multiple concepts. In diabetes, patient empowerment translates into improving adherence to agreed self-care regimens. Expected outcomes may include certain elements, for example weight loss, maintaining normal blood glucose levels, and the prevention of complications [22]. Chronic disease self-management programs, which strongly rely on
self-management education, have been tested in randomized controlled trials, showing significant improvements in the areas of healthy behaviors, communication with healthcare providers, health status and, most importantly, optimal use of healthcare [20].

Consistently with the available evidence, the International Charter of Rights and Responsibilities of People with Diabetes [18] aims at elevating the notion of patient empowerment to the status of a social right by according people with diabetes the right to information and education.

In particular, people with diabetes have the right to:

- receive the highest level of information and high-quality education about diabetes, including prevention and early detection strategies, particularly in high-risk individuals;
- gain knowledge on how the disease can be managed effectively and how to access education and clinical resources; be trained in self-management from the diagnosis onward, integrating the clinical, behavioral and psychosocial aspects of diabetes, either as a group or on an individual basis; be involved in assessing, planning, implementing, and routinely monitoring own care and health goals;
- access reliable information about the names and dosage of any relevant therapy/medication, including their action and potential side effects, as well as the interactions with other medical conditions and therapies, specific to the individual; share own information with care providers and access own medical records and other relevant information if explicitly requested.

On the other hand, the Charter states that people with diabetes have responsibility to:

- share information with their healthcare providers on their current state of health, which types of medicines they are using, any relevant allergies, social/lifestyle behaviors or other information that would be relevant for clinical decision making and expert advice;
- manage their agreed care and treatment plan;
- adopt, implement and monitor healthy lifestyle behaviors as part of their self-management of diabetes;
- share any problems they experience with their healthcare providers, including treatment plans and any barriers to their successful implementation;
- inform family, school, work and social colleagues of their disease condition, so that they can be supportive whenever needed;
- show consideration and respect for the rights of other people with diabetes and their healthcare providers.

The Charter considers people empowerment as a crucial aspect to combat the diabetes epidemic and its negative consequences.

Consistently with other aspects of modern society, the availability of timely information is crucial to control the effects of multiple factors under changing conditions. In the epidemiology of diabetes, it is important to simultaneously consider the impact of evolving scenarios in diverse areas that are not always immediately actionable, for example climate change, rapid change of environment, employment status, health system reform, healthcare settings, and socioeconomic determinants of health. Patient empowerment assigns to people with diabetes the task of actively improving the quality of their own healthcare, through difficult evaluation strategies that may include benchmarking own results against those obtained by other healthcare providers under similar conditions. Though compelling, this concept may not be easy to apply for all subjects.

The challenge extends to modern health information systems: empowered patients need comparable information, which is seldom available in a standardized format—particularly at the international level—and is rarely customized to respond to the needs of a specific category of users.

Diabetes represents an almost perfect model for analytical purposes, with its multiple dimensions and well-defined data structures. For this reason, diabetes has been defined as a data-driven disease which allows information to be effectively linked for multiple purposes [23]. During recent years, diabetes registers have been increasingly regarded as a comprehensive solution to collect reliable data and derive indicators of quality of care and outcomes, allowing an active use of health information by different categories of users [24].

The EU has closely followed these developments in recent policy statements. In 2006, the EU Conclusions on the promotion of healthy lifestyles and prevention of type 2 diabetes [25] urged Member States “to foster the collection, registration, monitoring and reporting at national level of comprehensive diabetes epidemiological and economic data as well as data on the underlying factors”; and to implement “evaluation systems with measurable targets to track health outcomes and cost-effectiveness, taking into account Member States’ organization and delivery of their respective health services …”.

At the same time, the European Commission committed to support the “networking and the exchange of information between Member States with a view to promoting best practice, reducing inequalities and optimizing healthcare resources; and examine and strengthen the comparability of diabetes epidemiological evidence by considering the establishment of standardized outputs for monitoring, surveillance and reporting of diabetes mortality, morbidity and risk factor data across Member States.”

The above concepts attained the highest degree of recognition in the EU Parliament resolution of 14 March 2012 on addressing the EU diabetes epidemic [26], which basically confirmed all the above statements on the necessity of a systematic data collection and monitoring of diabetes complications and health outcomes across Europe.

The implementation of efficient solutions for strengthening healthcare systems represents today a specific responsibility for policy makers and healthcare providers and a means to ensure the highest attainable level of health for people with diabetes. Through the different legislative instruments passed in recent years, and, above all, the International Charter of Rights and Responsibilities of People with Diabetes, patient empowerment and the right to health information have left the domain of
academic investigation to become an integral component of the societal strategy against the rising burden of diabetes.

**Conclusion: the way forward**

Nowadays, the role of people with diabetes should be considered as central in a complex interrelated network of political, social, and economic interests. The subtle implications of such a scenario need to be fully recognized and properly managed to improve the strategy to protect the right to health of people with or at risk of diabetes.

The diagram presented in Figure 76.1 represents the person with diabetes within a comprehensive perspective of mutual interaction.

People with diabetes stand at the top of a “pyramid of empowerment” composed of blocks of human actions, organized to respond to problems that originate from a series of connected domains (nested circles).

The foundations of the pyramid are laid by laws, regulations, and international charters which are continuously updated to support the life of people with diabetes within a changing environment, to ensure the respect of social values and human rights. The social and natural environment directly influences the adoption of public health measures, for example prevention policies and health promotion campaigns, whose content is based on the evidence produced by research, which in turn provides the basis for the preparation of clinical guidelines. The availability of guidelines ensures the provision of appropriate health services and regulates the market for medical devices and products, within a broader context of health systems that can be organized in different ways (e.g. universal coverage vs. private insurance). Effective solutions for the prevention of diabetes complications, for example integrated disease management, may be variously organized and supported by powerful technology.

As for many other aspects of modern society, being able to manage own rights strictly depends on the competent use of all information flowing across the different levels of the pyramid. Citizens must be able to evaluate their condition in a broader scenario, where each element contributes to the general conditions of social policy and healthcare.

Using medical records effectively is essential to close the loop of the empowerment cycle, but such capacity is limited by the ability of the system to integrate all sources and produce information that can be easily interpreted by anyone (e.g. health indicators). This requires the implementation of automated procedures that can provide access to updated population data in real time, through systematic and efficient health information systems.

In general, apart from basic indicators of limited interest (e.g. diabetes prevalence), the personal use of comparable health information is yet to be realized. However, recent developments in the field of disease registries, organized on top of routine records of episodes of care that are stored for different reasons (e.g. quality of care and expenditure control in diabetes, hypertension, and renal disease), could pave the way forward. Although the situation is rapidly improving, registries have often been developed only for well-resourced settings rather than entire populations, for which ad hoc surveys are still the solution of choice [27].

In a field such as diabetes, for example, where data is abundant but information is still scarce, population-based registers may represent a pragmatic solution connecting the interests of people with diabetes to those of public health intelligence [28]. Registers are population-based when they are structured in a way that results for a specific category of subjects (e.g. people with diabetes) can be compared against those of the general population, through denominators that are usually determined by storing a complete master index for a defined geographical area. This is evidently possible only if the public service itself, for obvious legal implications, is supervising the entire process, rather than the traditional organization of surveillance registries, for example cancer registries, where research institutions are at the heart of the entire process. Population-based registers, particularly those organized at a regional level, are usually intrinsic to the process of care and are aimed to deliver much more than the basics of producing public health indicators.

In the near future, modern forms of information delivery will serve the interest of people with diabetes more closely, through services such as robust and timely information on the epidemiology of diabetes and its complications; evaluation of
quality of care and interventions in a changing environment; estimates of the cost of the disease; interfaces for shared care and self-management; flags for high-risk subgroups of individuals; levels of adherence to treatment guidelines; data extraction for diabetes research.

To become a strategic element for the improvement of health outcomes [29], such information shall be open to everybody: national and local decision makers, healthcare administrators, health professionals, research institutions, and all citizens.

A practical example of a diabetes information system that operates consistently with the whole framework presented here has been realized by two sequential EU projects: “Best Information through Regional Outcomes” (BIRO) [30] and “European Best Information through Regional Outcomes in Diabetes” (EUBIROD) [31].

The BIRO [32] system implemented an open source architecture, based on the concept of “privacy by design,” which allows the cross-border flow of information across diabetes registers in Europe. The system is structured on a data model that regards each element (a regional/national government, the single care provider, or even a person with the disease) as a potential data source contributor. The same software, including advanced statistical routines, runs safely on the database stored at each node, using a predefined set of standardized criteria. Only aggregate data are transmitted to produce the pooled results. As the system is general, it delivers a wide range of quality and outcome diabetes indicators, compiled into a final common report template that can be used to inform all the above-mentioned categories of users.

The EUBIROD project delivered results obtained through the application of the BIRO information system in 20 countries, including demographics, clinical characteristics, risk factors, health system structures and processes, population rates, and risk-adjusted estimates of intermediate/terminal outcomes. EUBIROD also delivered an interface through which each data source can independently assess the level of adoption of privacy and data protection principles, including issues of data quality and access rights (“privacy performance assessment”) [33].

Novel approaches, for example those successfully applied in EUBIROD, undoubtedly show that there is enormous potential, at limited cost, still waiting to be exploited. Moving from prototypes to permanent systems will allow strengthening of the level of participation of the entire community, empowering each individual through direct access to relevant data and the independent assessment of trusted sources of information. This way, each person will be able to assess, discuss or dispute the accuracy of medical records, and use this information to set up eventual complaint procedures. The availability of all these elements will be equally crucial for ensuring the empowerment of people with diabetes.

Diabetes population-based registries, when based on an efficient information system, could represent a solid tool to ensure the right to health for people with diabetes: a right conceived as a comprehensive and multidimensional concept encompassing all relevant human and social rights, as well as the empowerment of people with diabetes.

References

30 BIRO project website: Available at: www.biro-project.eu/ (last accessed July 3, 2014).
31 EUBIROD website: Available at: www.eubirod.eu/ (last accessed July 3, 2014).
CHAPTER 77

The economics of diabetes care: a global perspective

Rhys Williams
Swansea University, Swansea, UK

Key points

• Increases in the prevalence and economic impact of diabetes have resulted in greater perceived importance and availability of good quality information on the economics of diabetes and diabetes care.
• Information is available which quantifies the substantial economic impact of diabetes on individuals, families, healthcare systems, and society. These descriptive studies have used a variety of methods.
• Analytical studies have identified interventions in primary prevention, early detection by screening and treatment of established diabetes which, when compared with current alternatives, are “cost saving” or “very cost effective.”
• Much of this evidence relates to developed countries despite the fact that 80% of the estimated 366 million people with diabetes live in low- and middle-income countries, which have particularly severe constraints on healthcare resources.
• Further work in this area needs to continue to inform clinicians and health policy makers so that scarce resources can be used efficiently to reduce the impact of diabetes and improve the quality of life of those who live with the condition.

The scope of this chapter

This chapter reviews some of the evidence that has accumulated, since the publication of the previous edition, on the economic impact of diabetes and on the cost effectiveness of interventions. It aims to provide clinicians and health policy makers with information on the economic impact of diabetes and an indication of how diabetes management and service development should evolve in relation to economic considerations. It does not claim to be a systematic review in the sense used, for example, by the Cochrane Collaboration but recently published studies have been selected if they are considered to contribute significantly to these aims. Given the need to restrict its length, the chapter cannot claim to be a comprehensive review of this burgeoning field.

Two literature searches (details available on request) were employed for this chapter. Search number 1 was highly sensitive with a wide range of economic terms. It yielded over 37,000 potentially relevant titles. Search number 2 was more specific and orientated more towards analytical studies. It yielded 850 potentially relevant titles. The remainder of the chapter is divided into two broad areas: studies of the economic impact of diabetes; and studies of the cost effectiveness or cost utility of interventions. Since the primary prevention of diabetes (or a significant delay in its onset) is the intervention likely to have the greatest reduction in its economic impact and since the evidence for the preventability of T2DM is now overwhelming, primary prevention is included in the term “diabetes care.”
The global literature now convincingly demonstrates that diabetes has considerable economic impact but that many interventions to prevent or delay its onset and to prevent or delay its complications are cost effective or even cost saving. However, the overwhelming majority of this evidence relates to developed (high-income) countries. A significant challenge now is to explore in much more detail the economic features of interventions in low- and middle-income countries—the location of much of the explosion in T2DM diabetes incidence and prevalence.

The economic impact of diabetes

Estimates of costs—from different perspectives

Fundamental to the quantification of the economic impact of any condition is the question of viewpoint, that is, when costs are being identified, is this from the point of view of the individual with the condition, that of the family, the health system, society as a whole, or a combination of these? The most comprehensive, recent information on diabetes healthcare expenditures is that published by the International Diabetes Federation (IDF) [3]. These data are estimates of healthcare expenditures by country which include spending by health systems as well as by individuals or families living with diabetes. The proportions attributable to systems on the one hand and to individuals on the other will vary from country to country according to the funding arrangements for the delivery of care.

Using formulae first put forward by Jönsson [4], the method used requires knowledge of the prevalence of diabetes, the total healthcare budget of each country, and an assumption of the ratio of the total cost of medical care for people with diabetes compared to the cost of care for people without diabetes — the “diabetes cost ratio” (R). These formulae enable the calculation of the total cost of care for people with diabetes (i.e., including costs both directly attributable to diabetes plus the costs of other care) and the costs specific to diabetes. Initial estimates of these costs [5] and the most recent estimates [3] have assumed likely lower and upper bounds of R of 2 and 3.

In 2011 an estimated USD 465 billion was spent worldwide in treating diabetes and preventing its complications [3]. Such a sum is difficult to comprehend in isolation. More meaningful, perhaps, is that this amounts to 11% of the world’s total health expenditure and that, per person with diabetes, the average annual expenditure is USD 1274. As will be emphasized later, this average figure conceals gross disparities in the resources available for diabetes care in different countries (and, quite plausibly, within countries). An estimated 20% of global diabetes expenditure is attributable to low- and middle-income countries in which 80% of people with diabetes live [3]. In high-income countries (such as the USA, UK, Japan, etc.) the average per person annual expenditure is USD 5063 compared with USD 271 in low- and middle-income countries. Conversion of these sums to International Dollars (ID) (thereby correcting for national differences in purchasing power) reduces this difference slightly—to ID 4888 compared with ID 456—but by no means eliminates it. Perhaps the most chastening figure of all is the estimate that, in 2011, the USA accounted for 43% (USD 201 billion) of the world’s expenditure on diabetes, this sum being spent on a mere 6.5% of the world’s diabetic community [3].

Estimating diabetes healthcare expenditure using such formulae is a “top-down” approach, that is, calculating expenditures on individuals using epidemiologic and financial data from populations. These IDF estimates do not differentiate between T1 and T2DM; however, for most countries, the bulk of these expenditures will be attributable to the latter. Differentiation of expenditures according to diabetes type is possible by this method given data for the prevalences of T1 and T2DM separately, but the diabetes cost ratio is likely to be significantly different for the two subtypes driven mainly by the different age profiles of the type 1 and type 2 populations. In most populations, hospital admission is the main driver of expenditures and it has long been known that, although the absolute incidence of hospital admission is greatest in the elderly, the greatest relative difference in admissions between individuals with diabetes and those without is seen in the young. Thus, the diabetes cost ratio for T1DM, in locations where diagnosis is possible and treatment is accessible, will be greater than that for T2DM.

A recent, detailed study which has used an alternative (and more commonly used) top-down approach to estimate T1 and T2DM costs separately is that of Hex et al. [6]. This used aggregated, routinely available data (on hospital admissions, prescriptions and primary care consultations, for example) and estimates (of premature mortality, for example) from the epidemiologic literature to quantify cost of T1 and T2DM separately in the UK in 2010/2011 and to predict what these costs are likely to be in 2035/2036. They considered both direct healthcare costs and indirect costs (i.e., the wider costs attributable to diabetes outside the realm of healthcare). In the former category were the costs of screening and testing (for diagnosis and retinopathy), those attributable to treatment and management of diabetes itself, and those attributed to treating its complications. Included in the indirect cost total were estimates for loss of productivity due to premature mortality, informal care, sickness absence (absenteeism) from work, and the potential loss of productivity among people with diabetes who remain at work (so called “presenteeism” [7]). The main features of their findings are summarized in Table 77.1. Even though these estimates relate specifically to the UK context, they illustrate a number of important features which will be applicable to other contexts.

First, Hex and coworkers’s estimate of indirect costs for both forms of diabetes combined exceeds those for direct costs (58.6% of costs vs. 41.4% in 2010/2011 and 57.5% vs. 42.5% in 2035/36) [6]. Thus, even though the economic impact of diabetes on healthcare is high, its impact outside the realm of healthcare is higher still. This finding in the literature is common though not universal (see, for example, reference [8]).
Table 77.1 Estimates of the economic burden of diabetes in the UK for 2010/2011 and projections to 2035/2036

<table>
<thead>
<tr>
<th>Costs</th>
<th>2010/2011</th>
<th>2035/2036</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All diabetes</td>
<td>Type 1 diabetes</td>
</tr>
<tr>
<td>Direct</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Screening &amp; testing</td>
<td>^a0.1%</td>
<td>^a0.2%</td>
</tr>
<tr>
<td>Treatment &amp; management</td>
<td>^c20.9%</td>
<td>^c28.7%</td>
</tr>
<tr>
<td>Complications</td>
<td>^c78.9%</td>
<td>^c71.1%</td>
</tr>
<tr>
<td>Total</td>
<td>^b£9.8 bn*</td>
<td>^b£1.0 bn</td>
</tr>
<tr>
<td>Indirect</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Premature mortality</td>
<td>^c34.4%</td>
<td>^c62.3%</td>
</tr>
<tr>
<td>Sickness absence</td>
<td>^c6.8%</td>
<td>^c10.5%</td>
</tr>
<tr>
<td>Presenteeism</td>
<td>^c21.9%</td>
<td>^c10.1%</td>
</tr>
<tr>
<td>Informal care</td>
<td>^c36.9%</td>
<td>^c17.1%</td>
</tr>
<tr>
<td>Total</td>
<td>^b£13.9 bn</td>
<td>^b£0.9 bn</td>
</tr>
<tr>
<td>Total</td>
<td>^b£23.7 bn</td>
<td>^b£1.9 bn</td>
</tr>
</tbody>
</table>

^aProportion of direct costs for that type of diabetes at that time.
^bProportion of all costs for that type of diabetes at that time.
^cProportion of indirect costs for that type of diabetes at that time.
^*£1 = USD 1.58
Source: Adapted from Hex 2012 [6].

Second, the predominant factor in direct costs is complications (via hospitalizations) accounting for between 70% and 80% of direct costs. Within this subcategory of complications (not shown in Table 77.1), the costs of renal failure and foot ulcers and amputations are the two largest clearly defined categories (although exceeded by “other cardiovascular disease”) accounting for, in 2010/2011, around one third of complication costs. Within this subcategory Hex et al. [6] have included an estimate for “excess inpatient days” on the assumption that a person with diabetes as a comorbid condition occupies a hospital bed for 2.6 days more than a similar person without diabetes for routine surgery [9]. This might be designated an “indirect, direct cost.”

In the category of indirect costs, the costs of loss of production from premature mortality predominate, accounting for, in 2010/2011, two thirds of indirect costs in T1DM and one third in T2DM. The costs of informal care are also substantial (17.1% and 38.3% of the 2010/2011 indirect costs in T1 and T2DM, respectively). Though not often costed in any detail, these aspects of care have long been known to be significant (e.g. see reference [10]) with contributions both in terms of the affected person and their relatives. Hex and coworkers’ estimates for “presenteeism” are equal, in proportional terms, to that for absenteeism in T1 but greatly exceed those for absenteeism in T2DM. The authors are clear that the methods for estimating the costs of presenteeism are in their infancy [6]. Nevertheless, these estimates are bound to be contentious since they imply that people with diabetes in work are considerably less productive than they should be, because of their diabetes.

Hex and coworkers’ predictions for costs in 2035/2036 make use of population projections and plausible assumptions about changes in prevalence and outcomes. Within the direct care category, the proportional contributions of the subcategories do not change greatly. However, within the indirect cost category, the proportional contributions of premature mortality costs are expected to fall (which is gratifying) whereas those of informal caring are expected to rise—which is more concerning. As might be expected, costs rise in all categories, driven by higher prevalences, higher treatment costs, and higher costs of complications. The total cost of diabetes (direct and indirect, type 1 and type 2) is predicted to rise by 68% on its 2010/2011 total while the rise for type 1 and type 2 separately are 121% and 63%, respectively [6]. The predicted increases in direct and indirect costs are roughly equal whereas the predicted increase of indirect costs for type 1 is, in proportional terms, the largest of all. The authors comment that “any improvements in the way diabetes is treated that lead to better glycemic control and fewer complications could have a significant impact on these costs, but this remains to be assessed.” To this should be added that reduction in prevalence through primary prevention would be expected to contribute significantly and that the goals of treatment should be widened to include other important parameters such as blood pressure and lipids. The authors’ comment about the need to take into account improvements in outcome is particularly pertinent given the observations being made in many high-income countries that the excess mortality resulting from diabetes is declining and that, as part of this, age-specific mortality from cardiovascular disease has decreased more in people with diabetes than in their nondiabetic peers. These changes and concomitant improvements in morbidities such as visual and renal impairment will need to be taken into account in revisions of models such as that of Hex et al. [6].
The bottom-up approach, that is, gathering data on individual expenditures and collating them to determine population-level costs, also enables estimates to be made specifically for T1 or T2DM without the need for diabetes cost ratio assumptions. A bottom-up approach is being used in the 3C Study [11] which aims to describe coverage, cost, and care of T1DM in the Beijing and Shantiou regions of China. The sampling unit is the individual with T1DM recruited from the outpatient and inpatient records of hospitals in the two regions. Interviews with participants, data from hospital billing departments, medical records departments, and government sources are being used in a combination of bottom-up and top-down approaches to calculate direct medical costs, direct nonmedical costs and indirect costs from different viewpoints—patients and their families, health system, insurers, and the wider society. This and other ongoing studies as yet unpublished may well confirm that the diabetes cost ratio, at least in some populations, is significantly greater than the 3 previously taken as the likely upper limit.

The evidence base for the costs of diabetes in countries other than the Americas, western Europe and Australasia is growing (see, for example, data from India [12,13], Ethiopia [14], and Iran [15]). However, Grover et al. [12] comment that “the data regarding cost of care of diabetes mellitus from developing countries are scarce.” A weakness of a number of these studies currently is that some are based on small samples of participants attending one (or a few) healthcare settings (e.g. [12,14]). Studies based on larger samples, preferably community based (e.g. [13,15]) are needed. Feleke et al. [14], however, and Esteghamati et al. [15] have the strength of nondiabetic participant data sampled from the same source.

Drivers of direct costs
Given that hospital admission for diabetes is an important driver of healthcare expenditures and that the bulk of hospital admissions of people with diabetes are for the treatment of complications, it follows that the presence of complications will be a potent influence on healthcare costs. Many studies have confirmed this.

The CODE-2 (Cost of Diabetes in Europe – Type 2) study [16] applied the same (bottom-up) methodology in eight European countries and demonstrated that the average healthcare cost per patient was 3.5 times greater in those with both micro- and macrovascular complications compared with those with no recorded complications. The presence of microvascular complications alone increased the per person costs by a factor of 1.7 and that of macrovascular complications alone doubled them. The presence of both types of complications increased costs specific to hospitalization by 5.5 times that found in the comparison group.

Since the publication of CODE-2, other studies have confirmed these findings. For example, the CODEIREE study (in the Republic of Ireland) [17] found that the presence of both micro- and macrovascular complications increased the annual cost of care by a factor of 3.8 compared to people with diabetes but with no complications; microvascular complications alone and macrovascular complications alone were associated with increases of 1.8 and 2.9, respectively. Morsanutto et al. [18] found that the presence of more than two diabetic complications increased the per person annual healthcare costs by a factor of 3 in their study of patients attending the diabetes centre of Portogruaro, northern Italy. More recently, Clarke et al. [19], using a different method, quantified the costs of diabetic complications in Western Australia. Their estimates for the costs of seven common complications—renal failure, lower limb amputation, heart failure, chronic leg ulcer, nonfatal stroke, ischemic heart disease, and nonfatal myocardial infarction—ranked in that order with the average costs in the year the event first occurred to a 60-year-old man ranging from AUD 28,661 (95% CI 22,989–34,202) for renal failure to AUD 11,660 (95% CI 10,931–12,450) for a nonfatal myocardial infarction (1 AUD ≈ 1.04 USD).

When individuals and families pay the cost of care themselves
The economic impact of diabetes can be particularly savage when much or most of it is borne by individuals or families. A larger share of healthcare expenditure for the condition is shouldered by individuals living in low- and middle-income countries when compared with those living in high-income countries [3]. In Latin America, for example, 40–60% of the cost of medical care for people with diabetes is out-of-pocket (i.e. financed from personal income) [20].

At least one study from India [21] has reported an increase over time in the extent to which families living with diabetes have to use their own income to fund diabetes care. Information from a sample of patients attending hospital clinics and general practitioners in urban and rural areas of seven states in India in 2005 showed that families in urban areas spent, on average, 10% of their income funding the care of a diabetic member while, for families in rural areas, the figure was 17%. Compared with a previous, similar study carried out in 1998 [22] and after accounting for inflation, the proportion of family income devoted to diabetes care had increased by 113%. In both studies, poorer families contributed larger proportions of their income than did the wealthier families and the secular increase was most marked in the poorest families increasing from 24.5% in 1998 to 34% in 2004. Sharing some of the characteristics of the poor in low- and middle-income countries are the uninsured and underinsured of the USA. A recent study of coverage for people with diabetes [23] compared health insurance coverage and type of coverage for 2704 adults with (self-reported) diabetes and 25,008 adults without, using data from the 2009 US National Health Interview Survey. Perhaps surprisingly, they found that coverage by any type of health insurance was higher (90%) in the group with diabetes than in those without (81%). More adults aged 18–64 years with diabetes had Medicare coverage compared to those without (14% vs. 4%), and more of those without diabetes
had private insurance (66% vs. 58%). Those within this age group who had diabetes were more likely to have more than one source of insurance coverage (13% vs. 5%). Songer [24] had commented earlier that 15% of families with one or more children reported difficulty in obtaining health insurance coverage because of pre-existing illness clauses. Casagrande and Cowie’s data [23] paints a more optimistic picture for adults.

Some of the consequences, at least in the process of care although not in outcome, of non- and under-insurance are detailed by DeVoe et al. [25]. Their findings, relating to 6562 people with diabetes aged over 18 years and with data (from 2002 to 2005) in the medical Expenditure Panel Survey, were similar to those of Casagrande and Cowie [23] in that more than 84% of people with diabetes had full-year coverage and a USC (usual source of care), and 2.3% had neither coverage nor a USC. However, being uninsured and without a USC was associated with a 5.5-fold higher likelihood of unmet medical needs (compared with being insured and having a USC) and a threefold likelihood of delayed urgent care.

The need for individuals and families in some populations to contribute towards the costs of care is exacerbated by curtailment in earning power as a result of diabetes. Songer [26], commenting on the work of Steen Carlsson et al. [27], has called for “the social consequences of type 1 diabetes [to] … be embraced more frequently in clinical settings.” The authors [27] had described similar earning power of people with diabetes and those without diabetes prior to its diagnosis but reduced earnings following diagnosis. With greater than 10 years diabetes duration and controlling for confounders, men with diabetes in employment and with upper secondary education were 4.2% below their nondiabetic peers in annual earnings and women 8.1% below. Songer [26] speculates that “with renewed enthusiasm for the long-term health of individuals after the diagnosis of diabetes [in more recent cohorts of patients], there is hope regarding the potential for full and productive employment.”

The financial vulnerability of people with diabetes has also been explored by Schofield et al. [28] and Holmes et al. [29] who have quantified the loss of income not only of people with diabetes but also of their carers. The former, Australian, study concentrated on the economic consequences of taking early retirement as a result of diabetes. The consequences were twofold: loss of income as a result of not being in employment, and reduced accumulated savings to finance their retirement as a result of their diabetes.

The findings of Holmes et al. [29] mirror those of other cost-of-illness studies that have found associations between healthcare costs and the presence of complications in that, although there was loss of earnings for both people with diabetes and their carers throughout the complications spectrum, this was most marked in instances when both micro- and macrocomplications were present. The burden of caring resulted in loss of earnings and significant levels of strain in the carers, as well as loss of earnings and significantly poorer health-related quality of life in the individuals being cared for.

**The consequences of less-than-optimum therapy**

The efficient use of resources is a topic for the next section of this chapter. The consequences of less than optimum therapy are, however, a useful link between descriptive economic studies and analytical studies such as cost-effectiveness and cost-utility analyses. Brown et al. [30] have examined the cumulative glycemic burden of treatment failure in diabetic members of the Kaiser Permanente Northwest Region Health Maintenance Organization concluding that, even in what was considered (by them) to be a well-controlled population, the average patient accumulated five HbA1c-years of excess glycemic burden compared with target levels and concluded from this that “clinicians should change glucose-lowering treatments in T2DM much sooner or use treatments that are less likely to fail.” More recently and also from the US, Nuckols et al. [31] have demonstrated that the cost of improving glucose control to goal (those of the Healthcare Effectiveness Data and Information Set (HEDIS)) being between USD 192 and USD 711 per person, is modest in relation to the extensive overall costs of providing care.

Several large studies have commented on what they consider to be less-than-optimal care. For example, the DiabCare Asia study audited the quality of care in 11 countries in 1998 and 2003 [32] and concluded, from the examination of the records of over 20,000 patients in the first and 15,500 in the second study, that the level of blood glucose control was, in their view, unsatisfactory (mean HbA1c ≈ 8% and mean fasting blood glucose ≈ 9 mmol L⁻¹). In the Bangladesh component of DiabCare Asia [33], similar conclusions were drawn with a mean HbA1c (amongst 1952 patients) of 8.6% with less than a quarter achieving the American Diabetes Association (ADA) target of <7%.

It is now an accepted tenet of diabetes care that poor metabolic control (not only of blood glucose but also blood pressure and lipids) is associated with an increased likelihood of the development of complications. It is also well known, from the preceding text and other evidence, that complications are associated with higher healthcare costs. If, as suggested by studies such as that of Nuckols et al. [31], improvements in control can be obtained at a modest increase in cost then it is tempting to posit that initiatives to improve control, if successful, will lead eventually to reduced overall costs. However, the difference in the time scale between the short-term increases in costs which are necessary to deliver improved quality of care and the longer time scale for the accrual of economic benefits resulting from the lower incidence of complications means that any reduction in overall costs may, at least in the short term, be elusive. For example, the International Diabetes Management Practice Study (IDMPS) [34], a 27-country initiative aiming to improve the quality of care through enhanced patient education, which significantly improved the proportion achieving metabolic targets, found an increase (though modest) in the cost of care. Any lower incidence of complications resulting from this improvement in care had not, at least at the time of publication, led to a reduction in overall costs.
This phenomenon is not confined to diabetes since the systematic review by Ofman et al. [35] which evaluated 102 studies of managed care for chronic disease (including 22 relating to diabetes) found few studies which convincingly showed a reduction in healthcare costs. This is probably because these initiatives usually lead to no or minimal improvements in clinical outcome or, when minimal improvements are achieved, the resultant cost savings are off-set by increases in the costs of healthcare delivery. The diabetes studies included are now rather dated since they span the years 1986 [36] and 2000 [37]. Amongst these 22 studies, only that of Naji et al. [38] includes a specific economic component though this focuses on additional patient costs, rather than global costs. The more recent systematic review of Knight et al. [39], which was specific to managed care in diabetes, included 24 studies (with considerable overlap with those included by Ofman et al. [35]) but their review concentrated on process and metabolic outcome measures rather than costs. Ofman et al. commented that “although disease management may improve the quality of life for patients with chronic disease [including diabetes], long-term studies may be required to show the economic benefits and financial return on investment.” The practicalities of empirical studies with sufficiently long-term follow-up may be such that the most feasible means of examining this question adequately will be modelling studies.

While it is logical to predict that initiatives to improve the quality of care may, in time, reduce the currently substantial costs of providing care to people with diabetes, it is misleading to propose that one of the aims of such initiatives should be to save money. In reality, healthcare resources which are freed up as a consequence of reductions in complications are not “saved” but will, inevitably and logically, be deployed for other purposes. It is not the aim of health systems and it is not the goal of health professionals to save money for the individual, the family, or the state. Rather, the goal is to improve the quality and increase the quantity of life for those affected by diabetes and other long-term conditions through the most effective and efficient use of resources.

**The cost effectiveness and cost utility of interventions**

Clinicians have a professional obligation to ensure that each patient they are treating receives effective care. They also have an obligation to their other patients that the treatment of that particular patient is as efficient as possible. Ethical clinical care thus requires consideration of both effectiveness and efficiency. Rational health policy making requires consideration of effectiveness, efficiency, and equity. Assessment of all three requires access to the best analytical evidence available.

**The cost-effectiveness plane**

The use of the cost-effectiveness plane as a diagrammatic illustration of the relationships between the costs and effects of interventions has been advocated by many, such as Black [40] and Laupacis et al. [41]. In this, two intersecting axes, one relating to cost, the other to effect (or outcome), form four quadrants (Figure 77.1). Quadrant A contains interventions which, in relation to the status quo (the intersection), deliver worse outcome and at higher cost. Clearly these are unacceptable in practice. In B, interventions deliver better outcome but at higher cost—the most commonly encountered case and one in which the acceptability of the intervention in practice will depend on its relative costs and benefits compared with alternatives. In C, interventions are cost saving—better outcome and at lower cost—clearly the most desirable of all while, in D, interventions are less costly but produce worse outcome. As emphasized by Morris et al. [42], the acceptability of interventions in B and D are worth considering in terms of their trade-offs; in B, the extent to which outcome is improved albeit at increased cost; and in D, the extent to which cost is reduced, albeit with worse outcome.

Morris et al. [42] divide zones B and D each into two sub-zones in which the trade-offs are of contrasting acceptability. Laupacis et al. [41] divide quadrant B into three such subzones (see Figure 77.1): (i) B1, in which interventions are likely to be regarded as cost effective (and, therefore, acceptable) because enhanced effects are achieved at relatively little extra cost; (ii) B2, in contrast, in which only small enhancement of effect requires comparatively large costs—interventions likely to be regarded as unacceptable; and (iii) B3, an intermediate zone in which the acceptability of the interventions will be the subject of debate about the trade-off. Li et al. [43], as described later, split quadrant B into four subzones.

Consideration of interventions falling in zone D raises some subtle questions of the balance between societal gain and individual loss. Particularly in situations where healthcare resources are severely limited, moving to less costly and less effective interventions may well be beneficial to the general population to an extent that mitigates the losses that some individuals will experience as a result of reduced effectiveness. For example, reducing the intensity and/or frequency of screening (perhaps for diabetic eye disease) or increasing the interval between scheduled clinic
visits for routine care will make available healthcare resources which can be used for other activities, either within or without the diabetes field. The resultant benefits of such changes to the general population may outweigh the losses that some individuals experience as a result of late detected retinopathy or other complications.

**Comparison of the acceptability of alternative interventions**

Economic comparisons of preventive and therapeutic alternatives are of three main types. Cost–benefit analyses identify the various costs of interventions and their potential benefits, assign each of them a monetary value and then compare total costs with total benefits to give a cost/benefit or benefit/cost ratio. The larger the latter (or the smaller the former) the more likely is the intervention to be acceptable. Cost–benefit analyses have the advantage of conceptual simplicity but suffer from the inherent difficulty of assigning valid monetary values particularly to benefits—extended life, reduction of disability or better well-being, for example.

Cost-effectiveness analyses express benefits in “natural units” of, for example, number of cases detected, percentage points of HbA1c reduction, or cases of disability prevented. Such analyses thus circumvent the need to ascribe monetary values to benefits but the main disadvantage of cost-effectiveness analysis is that interventions which lead to different types of benefits cannot be compared. Thus, for example, the cost effectiveness of an intervention aimed at preventing end-stage renal disease (ESRD) cannot be readily compared, using cost-effectiveness analysis, with an intervention aimed at preventing lower limb amputation. It can only be compared with interventions also aimed at ESRD.

Cost-utility analysis uses the concepts of QALYs (quality adjusted life years) or LYGs (life years gained) to provide common units for the comparison of the benefits derived from interventions aimed at different preventive or therapeutic outcomes. The QALY combines the notions of length of life and quality of life and requires assessments of “utility” which people would ascribe to different health states—dependency on renal dialysis versus limblessness, for example.

The incremental cost-effectiveness ratio (ICER) is often used to compare alternatives in cost-utility analysis. This is defined as the ratio between the difference in costs (those of the intervention minus those of an alternative) and the difference in effectiveness (or effects) between the two. The health economic literature (e.g. Morris et al. [42]) devotes considerable attention to the validity of this measure and the nature of alternative interventions for consideration. In practice, however, the usual alternative for comparison is either some form of usual care or treatment with a placebo. Since the aim of improving the effectiveness of interventions is to increase effectiveness at a minimum of increased costs, then the lower the ICER the better. ICERs are usually expressed in terms of cost per QALY or LYG.

**What do clinicians and policy makers need to know in terms of analytical studies in diabetes?**

It is clearly vital that clinicians and health policy makers have access to high-quality evidence as to which current and potential future interventions, within the epidemiologic and financial contexts in which they function, are cost saving or at least acceptably cost effective. Epidemiologic criteria will be important in respect of, for example, the pre-test probability of a positive result in the early detection of T2DM. This will be influenced by the prevalence of previously undiagnosed T2DM in the population. The local availability of resources will, in part, determine the boundaries between the different trade-off subzones of quadrants B and D in the cost-effectiveness plane.

The areas of practice in which these questions arise may be conveniently divided into three: (i) primary prevention of diabetes, (ii) its early detection (sometimes termed screening), and (iii) clinical management in established diabetes. Issues of cost effectiveness are not the only ones on which healthcare decisions are based. (There will be pragmatic, ethical or, particularly in the case of policy, political influences.) However, issues of cost effectiveness are becoming increasingly well recognized.

**What does the recent literature tell us about cost-saving and cost-effective interventions for primary prevention and for managing established diabetes?**

The recent, comprehensive review of Li et al. [43] considered interventions to prevent and control diabetes recommended by the ADA in 2008 [44]. English-language publications, from 1985 to May 2008 were reviewed following the protocol of the Cochrane Collaboration [45]. Studies were included only if they were considered “good” or “excellent” according to the BMJ authors’ guide for economic studies [46]. Interventions were judged according to their ICERS. The strength of evidence in each case was categorized as “strong,” “supportive,” or “uncertain” according to well-recognized criteria such as those of the Centre for Evidence-Based Medicine at the University of Oxford, and the ADA.

They set boundaries between four zones in quadrant B: (i) very cost effective: ICER $\leq$ USD 25,000 per QALY or LYG; (ii) cost effective: ICER USD 25,000–50,000; (iii) marginally cost effective: ICER USD 50,001–100,000; and (iv) not cost effective: ICER > USD 100,000. Of the original 9264 titles and abstracts identified, 56 were included in the final data analysis. They found 14 interventions for which strong evidence existed as being cost saving (6 interventions) or very cost effective (8 interventions). These are given in Table 77.2, categorized into the three areas: (i) primary prevention of diabetes (1 intervention); (ii) early detection of diabetes through screening (1); and (iii) clinical management in established diabetes (12).

It is important to note that all of the studies included were conducted in adults and all were based in high-income countries except one [47] from Bangkok, Thailand. This used
Table 77.2 Interventions with “strong evidence” of being cost saving or very cost effective according to Li et al. [43]. For details of primary references, see Table 77.1 of the original review

<table>
<thead>
<tr>
<th>Intervention</th>
<th>Comparison</th>
<th>Cost saving or very cost effective?</th>
<th>Country setting</th>
<th>Population(s) in which applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary prevention of diabetes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intensive lifestyle modification for the prevention of T2DM among people with impaired glucose tolerance</td>
<td>Standard lifestyle recommendations</td>
<td>Very cost effective</td>
<td>Australia, USA</td>
<td>Overweight and obese adults</td>
</tr>
<tr>
<td><strong>Early detection of diabetes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population-based screening for undiagnosed T2DM</td>
<td>No screening</td>
<td>Very cost effective</td>
<td>USA</td>
<td>African Americans aged 45–54 years</td>
</tr>
<tr>
<td><strong>Clinical management in established diabetes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACE-I therapy for the control of hypertension</td>
<td>Standard treatment for hypertension</td>
<td>Cost saving</td>
<td>UK, USA</td>
<td>People with early T2DM and hypertension</td>
</tr>
<tr>
<td>ACE-I or angiotensin receptor blocker (ARB) therapy in the prevention of end-stage renal disease</td>
<td>No treatment by ACE-I or ARB</td>
<td>Cost saving</td>
<td>US</td>
<td>People with T2DM and hypertension initially free of CVD and ESRD</td>
</tr>
<tr>
<td>Early irbesartan therapy following the advent of microalbuminuria to prevent ESRD</td>
<td>Treatment delayed until macroalbuminuria present</td>
<td>Cost saving</td>
<td>Canada, UK, USA</td>
<td>People with T1 or T2DM and microalbuminuria</td>
</tr>
<tr>
<td>Comprehensive foot care to prevent ulceration</td>
<td>Usual care</td>
<td>Cost saving</td>
<td>Sweden</td>
<td>People with T1 or T2DM</td>
</tr>
<tr>
<td>Multicomponent intervention* targeted at risk factor control and early detection of complications</td>
<td>Conventional care</td>
<td>Cost saving</td>
<td>Germany, Switzerland, UK, USA</td>
<td>People with T1DM</td>
</tr>
<tr>
<td>Multicomponent intervention* targeted at risk factor control and early detection of complications</td>
<td>Conventional care</td>
<td>Cost saving</td>
<td>Germany, Switzerland, UK, USA</td>
<td>People with T2DM</td>
</tr>
<tr>
<td>Intensive glycemic control</td>
<td>Standard treatment</td>
<td>Very cost effective**</td>
<td>Japan, Switzerland, UK, USA</td>
<td>People with newly diagnosed T2DM</td>
</tr>
<tr>
<td>Statin therapy for secondary prevention of cardiovascular disease</td>
<td>No statin therapy</td>
<td>Very cost effective</td>
<td>Canada, Ireland, UK, USA</td>
<td>People with T2DM and CVD history</td>
</tr>
<tr>
<td>Smoking cessation counselling and treatment</td>
<td>Neither</td>
<td>Very cost effective</td>
<td>USA</td>
<td>People with newly diagnosed T2DM who are smokers</td>
</tr>
<tr>
<td>Annual screening for retinopathy and ensuing laser treatment</td>
<td>No screening</td>
<td>Very cost effective</td>
<td>Switzerland, USA</td>
<td>People with T1DM</td>
</tr>
<tr>
<td>Annual screening for retinopathy and ensuing laser treatment</td>
<td>No screening</td>
<td>Very cost effective</td>
<td>Switzerland, USA</td>
<td>People with T2DM</td>
</tr>
<tr>
<td>Immediate vitrectomy for the treatment of retinopathy</td>
<td>Deferred vitrectomy</td>
<td>Very cost effective</td>
<td>USA</td>
<td>People with T1 or T2DM and retinopathy</td>
</tr>
</tbody>
</table>

*Multicomponent intervention, standard antidiabetic care plus education, nephropathy screening, ACE-I treatment and retinopathy screening.

**Defined in terms of microvascular disease outcomes of the results of studies published prior to the ADDITION study. Results from the latter [69] make this conclusion less secure.

Source: Adapted from Li 2010 [43].
effectiveness data derived from developed countries but, using cost data and survival data from their own country, the authors were able to describe, as cost saving, the use of angiotensin-converting enzyme inhibitors (ACE-I) in slowing the progress of nephropathy in normotensive patients with T2DM and microalbuminuria.

Li et al. [43] commented that “a large majority of the ADA recommended interventions are cost-effective.” Their review provides many answers but also raises questions: are the interventions listed in Table 77.2 cost saving or very cost effective in other contexts? Will better evidence recruit more interventions to these categories? Can the efficiency be improved of interventions which they categorize as merely cost effective or marginally cost effective? Those most likely to be acceptable in low- and middle-income countries might be: intensive lifestyle modification for the prevention of T2DM (though this would depend on the degree of intensity required); population-based screening for undiagnosed T2DM in certain ethnic groups within appropriate age bands; comprehensive foot care to prevent ulceration; statin therapy for secondary prevention of cardiovascular disease; smoking cessation counselling and treatment (though this would depend upon the availability of appropriate supportive therapies). The world literature clearly requires periodic assessments of this type but the more immediate need now is for more primary research and subsequent systematic review of interventions which are feasible in low- and middle-income settings.

Narayan and Wilkinson [48], while acknowledging the effectiveness in randomized clinical trials (RCTs) of lifestyle interventions, emphasize the different economic context of prevention programs based in communities in which a large number of participants will be normoglycemic—in whom interventions may be likely to have little impact on future diabetes incidence. Herman [49] has also examined this area, summarizing studies of the cost effectiveness of primary prevention of T2DM—from the USA [50,51], Canada [52], Australia [53], Germany [54], one study from the viewpoints of Australia, France, Germany, Switzerland, and the UK [55], and one from Spain, Germany, and Sweden [56]. All but three of these were included in the review of Li et al. [43]. The exceptions were Herman et al. [50], Icks et al. [54], and Josse et al. [56]. Herman’s conclusions [50] are qualitatively similar to those of Li et al. [43], namely, that lifestyle modification as a primary prevention intervention for T2DM is either cost saving [53,55] or extremely cost effective [52,54]. Further, that three of the four published studies that had included metformin in this regard had drawn the same conclusion (the exception being Eddy et al. [51]) and that the two studies including acarbose concluded either that it was cost saving [52] or extremely cost effective [56].

The two reviews cited above and the primary research they summarize all relate to developed countries, settings in which the numbers of people with previously undiagnosed T2DM or prediabetes (impaired glucose regulation outside the diabetic range) although significant, are likely to be lower than in, for example, the countries of Asia or communities of Asian origin in countries elsewhere. As in developed countries, the within-trial cost effectiveness of interventions for the prevention of T2DM in some of these populations is well established (see, for example, Ramachandran et al. [57]) with lifestyle modification again being the more cost effective compared to metformin. Community-based prevention programmes in populations with a high prevalence of undiagnosed disorders of glucose regulation are likely to be even more cost effective than in populations in, for example, the USA and Western Europe, partly because the costs per case detected will be lower.

In a recent review of all aspects of diabetes in Asian populations, Ramachandran et al. [58] comment that “more data are needed for the economics of diabetes and quality of life and cost-effectiveness of various interventions [in Asian populations].” Particularly valuable would be long-term economic information relevant to community-based primary prevention programmes focusing on whether and to what extent interventions proven to be effective and cost effective in RCTs are cost effective (or even cost saving) in less than ideal settings. While information from empirical studies would be preferable, well-conducted modelling studies would be the “next best thing” given the long-term nature of follow-up and the considerable sample sizes that would be required. Well-conducted modelling studies would maximize the sources of data derived from ethically and socially appropriate empirical sources.

Few aspects of diabetes care generate quite as much controversy as the early detection of (or screening for) T2DM. As listed in Table 77.2, opportunistic screening is regarded as very cost effective in certain high-risk groups. However, for lower risk populations, the current consensus is that it is cost effective as part of a concerted programme for vascular risk assessment, risk reduction, and risk management including both the early detection of both T2DM and lesser degrees of glucose intolerance such as IGT [59,60]. Particularly topical is the question of HbA1c versus blood glucose as the screening test. Though the former is the more costly, its greater predictive value in relation to future complications may make it a cost-effective choice [61].

The information that we have is not what we need most

In 1971 the general practitioner, Julian Tudor Hart, described the “Inverse Care Law” the first part of which is often quoted: “The availability of good medical care tends to vary inversely with the need of the population served” [62]. The currently available literature on the cost effectiveness of interventions to prevent and control diabetes, as illustrated earlier, relates, in the main, to developed countries, yet 80% of the estimated 366 million individuals with diagnosed or undiagnosed diabetes in the world live in low- and middle-income countries [3]. Available knowledge on this important topic, therefore, may be said to illustrate a “law” analogous to the inverse care law—the
Table 77.3 Low-income countries, USA, and key diabetes-related data

<table>
<thead>
<tr>
<th>Country</th>
<th>Diabetes cases (20–79) in 1000s</th>
<th>Diabetes-related expenditure*</th>
<th>IGT national prevalence (%)</th>
<th>Country</th>
<th>Diabetes cases (20–79) in 1000s</th>
<th>Diabetes-related expenditure*</th>
<th>IGT national prevalence (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afghanistan</td>
<td>818.30</td>
<td>81</td>
<td>5.97</td>
<td>Kyrgyz Republic</td>
<td>154.23</td>
<td>79</td>
<td>4.69</td>
</tr>
<tr>
<td>Bangladesh</td>
<td>8,405.61</td>
<td>27</td>
<td>2.44</td>
<td>Liberia</td>
<td>51.02</td>
<td>47</td>
<td>11.08</td>
</tr>
<tr>
<td>Benin</td>
<td>70.63</td>
<td>60</td>
<td>10.65</td>
<td>Madagascar</td>
<td>428.01</td>
<td>38</td>
<td>9.53</td>
</tr>
<tr>
<td>Burkina Faso</td>
<td>175.13</td>
<td>69</td>
<td>8.93</td>
<td>Malawi</td>
<td>352.26</td>
<td>31</td>
<td>8.42</td>
</tr>
<tr>
<td>Burundi</td>
<td>94.22</td>
<td>35</td>
<td>7.49</td>
<td>Mali</td>
<td>99.83</td>
<td>70</td>
<td>9.82</td>
</tr>
<tr>
<td>Cambodia</td>
<td>199.37</td>
<td>82</td>
<td>8.91</td>
<td>Mozambique</td>
<td>294.72</td>
<td>37</td>
<td>10.46</td>
</tr>
<tr>
<td>Central African Rep.</td>
<td>58.16</td>
<td>34</td>
<td>10.41</td>
<td>Myanmar</td>
<td>2,103.67</td>
<td>16</td>
<td>8.40</td>
</tr>
<tr>
<td>Chad</td>
<td>197.05</td>
<td>86</td>
<td>9.21</td>
<td>Nepal</td>
<td>488.20</td>
<td>44</td>
<td>1.96</td>
</tr>
<tr>
<td>Comoros</td>
<td>23.48</td>
<td>46</td>
<td>9.18</td>
<td>Niger</td>
<td>283.85</td>
<td>39</td>
<td>8.19</td>
</tr>
<tr>
<td>Congo, Dem. Rep.</td>
<td>94.61</td>
<td>143</td>
<td>7.46</td>
<td>Rwanda</td>
<td>126.23</td>
<td>82</td>
<td>8.17</td>
</tr>
<tr>
<td>Eritrea</td>
<td>95.49</td>
<td>17</td>
<td>8.39</td>
<td>Sierra Leone</td>
<td>72.07</td>
<td>83</td>
<td>10.17</td>
</tr>
<tr>
<td>Ethiopia</td>
<td>1,377.28</td>
<td>25</td>
<td>8.18</td>
<td>Somalia</td>
<td>185.14</td>
<td>20</td>
<td>10.25</td>
</tr>
<tr>
<td>Gambia, The</td>
<td>12.29</td>
<td>46</td>
<td>11.85</td>
<td>Tajikistan</td>
<td>155.45</td>
<td>61</td>
<td>4.30</td>
</tr>
<tr>
<td>Guinea</td>
<td>181.79</td>
<td>37</td>
<td>10.03</td>
<td>Tanzania</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Guinea-Bisau</td>
<td>19.13</td>
<td>31</td>
<td>9.54</td>
<td>Togo</td>
<td>81.16</td>
<td>67</td>
<td>10.77</td>
</tr>
<tr>
<td>Haiti</td>
<td>295.46</td>
<td>68</td>
<td>3.95</td>
<td>Uganda</td>
<td>307.91</td>
<td>84</td>
<td>7.59</td>
</tr>
<tr>
<td>Kenya</td>
<td>769.29</td>
<td>57</td>
<td>9.36</td>
<td>Zimbabwe</td>
<td>550.86</td>
<td>56</td>
<td>5.70</td>
</tr>
<tr>
<td>Korea, Dem. Rep.</td>
<td>1,507.50</td>
<td>17</td>
<td>10.83</td>
<td>USA</td>
<td>23,721.77</td>
<td>8,468</td>
<td>11.97</td>
</tr>
</tbody>
</table>

*Mean diabetes-related expenditure per person with diabetes (USD)

Source: International Diabetes Federation 2011 [3].

“law of the inverse availability of evidence.” For prevention and control of diabetes in the majority of the world’s populations we have, as yet, insufficient evidence on cost effectiveness. This is ironic given the likelihood that, in these countries, the returns on investment both in economic and human terms (quality of life gained and reduced years of life lost) are likely to be greater per dollar, rupee, pound, peseta, euro, etc. spent than in high-income countries. This follows from the size of the populations concerned, their likely high prevalence of undiagnosed diabetes and their likely large numbers of relatively poorly controlled individuals for whom improvements in control have greater benefits than already close to accepted targets.

The World Bank classifies national economies according to Gross National Income (GNI) per capita [63]. In 2010 prices, the four bands are: (i) high income: USD 12,276 per capita or more; (ii) upper middle income: USD 3,976 to USD 12,275; (iii) lower middle income: USD 1,006 to USD 3,975; (iv) low income: USD 1005 or less. Low-income countries, as judged by these criteria, are listed in Table 77.3. Also illustrated in Table 77.3 is the number of people aged 20–74 years estimated to have diabetes, the mean annual diabetes-related expenditure per person and the national prevalence of IGT for each country and for the USA, the country from which most of the cost effectiveness evidence reviewed above is derived.

The number of people with diabetes in these countries for which estimates are available is around 18.5 million. Nine of them have prevalences of IGT comparable with that of the USA. Considerably more is spent per year on each person with diabetes in the USA (USD 8468) than is spent in all of the low-income countries combined (USD 1815) and, as a further measure of inequality, the US expenditure per person per year is 500 times that of the poorest countries (Eritrea, Democratic Republic of Korea, and Myanmar). The IDF Diabetes Atlas [3], from which these data are derived, comments that Luxembourg (not shown in Table 77.3) spends USD 9341 per year for every person with diabetes compared to less than USD 20 in those three poorest countries.

The 37,000 potentially relevant titles identified by search number 1 contained very little information of direct relevance to these low-income countries. An exception is that of Bangladesh which, amongst other topics, has information relevant to the early detection of diabetic neuropathy [64] and nephropathy [65], and the management of the diabetic foot [66]. Given some caution, information from such studies can be extrapolated to other similar contexts. Results from RCTs conducted elsewhere may be combined with local epidemiologic and financial data [47] in modelling studies or, as Baik et al. [67] have observed, locally generated observational data may, under certain circumstances, be of considerable value in providing clinicians and policy makers with information in the absence of locally conducted analytical studies.

As Mbanya [68] commented, in the previous edition of this textbook: “we need studies to explore the feasibility, effectiveness, and cost effectiveness of interventions, particularly brief interventions at primary care settings and of community-wide [preventive] interventions.” His remarks were made in the context of sub-Saharan Africa but they are applicable to all developing countries. There have been notable achievements since these words were published but there is still a considerable way to go.
Acknowledgments

I am grateful to Katrina Dalziel and Sue Prosser for assistance with searching the literature and the retrieval of articles.

References

American Academy of Clinical Endocrinologists
American College of Cardiology/American Heart Association (AHA/ACC) guidelines,
American Heart Association/American College of Cardiology/American Heart Association (AHA/ACC) guidelines,
American College of Cardiology/American Heart Association (AHA/ACC) guidelines,
American College of Cardiology/American Heart Association (AHA/ACC) guidelines,
American College of Cardiology/American Heart Association (AHA/ACC) guidelines,
American College of Cardiology/American Heart Association (AHA/ACC) guidelines,
American College of Cardiology/American Heart Association (AHA/ACC) guidelines,
American College of Cardiology/American Heart Association (AHA/ACC) guidelines,
American College of Cardiology/American Heart Association (AHA/ACC) guidelines,
American College of Cardiology/American Heart Association (AHA/ACC) guidelines,
American College of Cardiology/American Heart Association (AHA/ACC) guidelines,
American College of Cardiology/American Heart Association (AHA/ACC) guidelines,
American College of Cardiology/American Heart Association (AHA/ACC) guidelines,
American College of Cardiology/American Heart Association (AHA/ACC) guidelines,
American College of Cardiology/American Heart Association (AHA/ACC) guidelines,
American College of Cardiology/American Heart Association (AHA/ACC) guidelines,
American College of Cardiology/American Heart Association (AHA/ACC) guidelines,
American College of Cardiology/American Heart Association (AHA/ACC) guidelines,
American College of Cardiology/American Heart Association (AHA/ACC) guidelines,
American College of Cardiology/American Heart Association (AHA/ACC) guidelines,
American College of Cardiology/American Heart Association (AHA/ACC) guidelines,
American College of Cardiology/American Heart Association (AHA/ACC) guidelines,
American College of Cardiology/American Heart Association (AHA/ACC) guidelines,
American College of Cardiology/American Heart Association (AHA/ACC) guidelines,
American College of Cardiology/American Heart Association (AHA/ACC) guidelines,
beta-cells (β-cells) (continued)

- mass, 116–17
- in T2DM, 103, 354–5, 357, 358, 361
- meglitinide effects, 616–18
- metabolism, 96–9
- oxidative phosphorylation, 97, 420
- origin of new cells in adult, 64
- pathophysiology, 103
- regeneration, 73, 356–8
- regulation of growth, 64–5
- sulfonylurea effects, 616–18
- in transplantation of pancreas or islets, function, 777

β-granule, 89, 91, 92, 101

- exocytosis/insulin secretion, 99, 100–1
- abnormalities, 103, 364
- β-oxidation

- compared with
- combined with two other drugs, 683, 696–7
- adverse events, 650–1, 679
- clinical applications
- microvascular complication prevention, 646–8, 877
- nonalcoholic fatty liver disease, 298, 649–50
- polycystic ovary syndrome, 649
- pregnancy, 648, 830
- T2DM, 645–8, 820–1, 877
- T2DM prevention, 554–6, 648
- weight loss properties, 500, 648–9
- combined with one other drug, 645–6
- α-glucosidase inhibitors, 675–6
- dipeptidyl peptidase-4 (DPP-4) inhibitors, 647, 690, 692–3, 698
- insulin, 697
- meglitinides, 690–1, 694–5
- in poorly-controlled T2DM, 694
- PPAR-γ agonists (thiazolidinediones), 660, 661, 689, 691–2, 697, 698
- sodium–glucose transport-2 inhibitors, 693
- sulfonylureas, 689–91, 693, 694–5
- combined with two other drugs, 683, 696–7
- compared with α-glucosidase inhibitors, 675
- contraindication and adverse side-effects, 650–1
- metabolic effects, 662–4
- obese adolescent T2DM, 820–1
- pharmacology, 642–5
- plasmogen activator inhibitor 1 and effects of, 648, 1058
- product formulations, 645–6
- bile acids
- bariatric surgery effects, 513
- sequastrents, 269, 758–61
- biliopancreatic diversion with or without duodenal switch (BPD-DS; BPD), 505, 506, 511, 512
- bariatric surgery effects, 511
- glycemic control, 510
- mechanism, 507–8
- binge-eating disorder, 848, 849
- bio-artificial pancreas, 771
- biochemistry see laboratory values/tests; metabolism biomarker monitoring, 853–71
- biopsy
- nerve, 926–7
- skin
- in neuropathy, 932
- in Rosenbloom syndrome, 959
- Biosstat, 745
- BIRCO (“Best Information through Regional Outcomes”), 1111
- bisphosphonates, Charcot neuroarthropathy, 972
- bivalirudin thromboprophylaxis, 1059
- BL1282, 711
- black Americans as African Americans
- bladder dysfunction, 948
- blindness (visual loss) threat with retinopathy, 890, 891
- blood flow, myocardial, thiazolidinedione effects, 664–5
- blood pressure (arterial)
- abnormal see hypertension; hypotension
- α-glucosidase inhibitor effects, 678
- ambulatory monitoring, 921
- diurnal variations, 1010
- goal/target, 1083
- renal disease and, 566–7, 920–1, 1008, 1085
- end-stage, 913
- thiazolidinedione effects on, 665
- blood pressure control/lowering (incl. antihypertensive drugs), 1083–6
- in nephropathy prevention, 566–7, 920–1
- in neuropathy prevention, 568–70
- in retinopathy management, 898
- Blue Mountains Eye Study, 890
- body mass index (BMI)
- bariatric surgery and, 509
- nonalcoholic fatty liver disease, 286
- obstructive sleep apnea and, 319, 320–2, 324, 325–6, 328
- bombesin-like peptides, 127
- bone health
- adverse effects on, 953–4
- high-protein low-carb diet effects on, 494
- thiazolidinedione adverse effects on, 669
- boron affinity chromatography in HbA1c measurement, Hb variants affecting, 862, 863
- bottom-up approach to estimating costs, 1116
- brain (incl. cerebrum and central nervous system), 388
- edema complicating diabetic ketoadipsis, 812
- energy balance and, 529–37
- functional MRI see functional MRI
- GLP-1 effects, 151, 727
- glucose control and, 529–37
- response to glucose shortage, 787–8
- glucose shortage in (glycemia), 785–6
- peripheral neuropathy effects on, 965
- T2DM and, 388
- brain atrial natriuretic (BNP) and insulin resistance, 345, 346
- branched-chain amino acids, 278
- breads
- glycemic index, 582
- optimal amounts, 584
- breakfast cereals
- glycemic index, 582
- optimal amounts, 584
- breast cancer, 306, 307, 308, 309, 310, 312, 313
- breastfeeding, 833
- bromocriptine-QR, 761–2
- BTP57582, 711
- buformin, 298
- bulimia, 848
- bupropion, weight loss properties, 502
- burden (impact) of diabetes
- cardiovascular disease, 1031–2
- economic see cost
- hypoglycemia, 790
- St Vincent declaration on reducing, 1108

Bypass Angioplasty Revascularization Investigation in Diabetes (BARI) trials, 1058, 1074
- bypass graft, coronary artery, 1074

C

- C-cell carcinoma and GLP-1 agonists, 738
- C-reactive protein, 44
- insulin resistance and, 344
- Caenorhabditis elegans model, 519–20
- calcitonin gene-related polypeptide (CGRP), 131
- calcitonin receptor and calcitonin-receptor-like receptor, 131
- calcium (ions)
- hypertension and, 1080, 1081
- insulin secretion and, 99
- calcium channel blockers (CCBs) in hypertension, 1085, 1086
- caloric (energy) content and intake of diet, 578–9
- high, and cancer, 312
- managing/restricting/reducing, 490–3
- current recommendation in diabetes, 578, 578–9
- in nonalcoholic fatty liver disease, 293–4
- in pregnancy complicated by diabetes, 829
- calorimetry
- direct, 479–80
- indirect, 480
- Canadian – European Cohort Study Group, 542
- see also United States and North America canakinumab, 543
- cancer (malignancy)
- colorectal see colon and colorectal cancer
- insulin therapy and risk of, 795
- metformin effects, 650
- pancreatic see pancreas
- T2DM and, 306–15
- thiazolidinedione effects, 666–7
- thyroid, and GLP-1 agonists, 738
- Cancer Prevention Study II (CPS II), 306, 308
- candesartan, retinopathy, 898
- CAP-Ch1 complex, 172–4
- capillary electrophoresis – mass spectroscopy
- nephropathy detection, 913–14
- carmamylated hemoglobin, 863
- carbohydrate
- dietary, 44
- current recommendation in diabetes, 578, 580–1
- diet loss in see low-carbohydrate diets
- inhibitors of intestinal digestion and absorption of, 710
- in nonalcoholic fatty liver disease management, 293–4
- see also sugar
- metabolism, ketosiasis and, 801–2
- carbohydrate responsive element binding protein (ChREBP), 207, 264
- carboxypeptidase E (CP-E), 533
- carboxypeptidase H (CP-H), 89–91
- CAREDia (Coronary Artery Revascularization in Diabetes) study, 1074
- cardiomyopathy, 941–2, 1068–9
- cardiovascular disease (CVD), 1007–1011
- burden, 1031–2
- deaths see mortalities
- diabetes therapy and its impact on risk of, 1070–3
- dyslipidemia and, 262–3
- endothelium and see endothelium
- erectile dysfunction and risk of, 980, 982
- HbA1c and risk of, 859–60
- hyperglycemia and see hyperglycemia
- insulin resistance and, 346
- nephropathy and risk of, 923
- nonalcoholic fatty liver disease and risk of, 289
obstructive sleep apnea and, 329–40
older adults, risk factor reduction, 839
periodontal disease and, 907
transcription factor MODY and, 432
see also heart; macrovascular disease
cardiovascular drugs associated with erectile dysfunction, 979
cardiovascular system
autoimmune neuropathy of, 941–4, 1067–8
drugs and their effects on α-glucosidase inhibitors, 678
DPP-4 inhibitors, 737
GLP-1 receptor agonists, 737
sulfonilureas, 631
thiazolidinediones, 664–5
GLP-1 effects, 152, 727
hypoglycemia adverse effects, 790
CARDS (Collaborative Atorvastatin Diabetes Study) trial, 265, 267, 268, 922
CARE (Cholesterol and Recurrent Events) trial, 268
care see management
cardiotonic arteries
atherosclerosis, 1011
tima – media thickness and effects of α-glucosidase inhibitors, 678
revascularization, 1098
carpal tunnel syndrome, 955, 958
casts
Charcot neuroarthropathy, 972
foot ulcers, 969
cataracts, 902
surgery, 902
retinopathy and, 892
catecholamines and β-cells, 217
classification see classification and categorization
cation-exchange chromatography see ion-exchange chromatography
causes/etiology (incl. predisposing/risk factors) in classification of diabetes, 4, 5
coronary heart disease, 1013–14, 1017
gene see genetic defects and factors hyperinsomolar state, 801, 805–6
hypoglycemia (severe), 789
assessing risk, 791
ketoadiposis, 801
nephropathy, 912–13
nonalcoholic fatty liver disease, 286–8
obesity, energy expenditure in, 482–3
obstructive sleep apnea, 317–18
peripheral disease as risk factor, 995–6, 1001
retinopathy, 892
stroke, 1016–17
T1DM, 18–23
T2DM, 7
cancer and, 308
children, 41–2
modifiable, 42–5
see also pathogenesis
cavernosal tissue see corpus cavernosum
Cbl1-CAP complex, 172–4
CLL2 and CCL5, 885
CD2-targeted drug in T1DM, 543
CD5 monoclonal antibody therapy, T1DM, 542
CD4+ T-cells, multiple types and autoreactive, 447
CD8+ T-cells, multiple types and autoreactive, 447
CD20 monoclonal antibody therapy, T1DM, 542
CD25 (interleukin-2 receptor α chain) and its gene, 443, 446, 460, 462
CD34+ endothelial progenitor cells and cardiovascular events, 1047
CD36 (fatty acid translocase), 524, 525, 880, 1055, 1069
CD4 see Centers for Disease Control and Prevention
cell cytoskeleton, GLUT4 translocation and, 178
differentiation (cytodifferentiation), 59–62
cell-based therapies, 62
T1DM, 543–4
see also stem cell therapy
cell cycle regulators of β-cell growth, 64–5
cellular mechanisms
cardiomyopathy, 1069
hyperglycemia effect on insulin secretion, 420–1
hypertension, 1080–1
insulin actions, 225–6
insulin resistance in T2DM, 381–5
Centers for Disease Control and Prevention (CDC) and American Academy for Periodontology (joint group), 989, 992, 993, 994
T2DM, 449
Central America, T2DM age and sex-specific prevalence, 31–2
central fat see abdominal fat
central immune tolerance, 445
central nervous system see brain (incl. cerebrum and central nervous system)
central sleep apnea (CSA), 316, 331
obstructive see obstructive sleep apnea
ceramide and atherogenesis, 1039
cereals (grains)
glycemic index, 582
infant diet, 22
optimal amounts, 584
see also breakfast cereals
cerebral... see brain
cerebrovascular disease, 1015–17, 1097–9
atherosclerosis, 1011
epidemiology, 1015–16
cesarean vs. vaginal delivery, 832–3
Charcot neuroarthropathy, 972–3
charge difference method of HbA1c measurement, 861
Charter of Fundamental Rights (EU), 1107
chemical-induced diabetes, 8–9
chelation Co-C-Motif Ligand-2 and 5 (CCL2 and CCL5), 1020
Chicago Heart Association Detection Project in Industry Study, 1018
CHICAGO Heart study, 1012, 1018
CHICAGO Heart Association Study, 1012, 1018
neonatal diabetes, 434
permanent, 434
temporary, 435
obstructive sleep apnea, 318–19
Wolfman syndrome, 437
clinical trials/studies, 539–73
insulin pumps, 748–9, 751
new drugs, 709, 710
T1DM
preventive, 544–6
therapeutic, 541–4
T2DM, preventive, 550–63
clinician see physician
Cock, 1057, 1058
clonidine, 847
clonidine, 847
CoA syntheses, 847
coenzyme Q10, 884
in maternally inherited diabetes and deafness, 437
cognitive – behavioral therapy (CBT)
anxiety, 847, 848
depression, 847
cognitive—behavioral therapy (CBT) (continued)
eating disorders, 848, 848–9
cognitive (neocognitive) impairment, 849
elderly, 839–40
colesavelam, 269, 758–61
colestipol, 269
Collaborative Atenostatin Diabetes Study (CARDS) trial, 265, 267, 268, 922
collaborative management see multidisciplinary team
collagen (skin) and Rosenbloom syndrome, 960
congenital malformations with diabetes in pregnancy, 960
colon and colorectal cancer α-glucosidase inhibitors and, 680
T2DM and, 307, 309, 310, 311–12
colony-stimulating factor, granulocyte see granulocyte-colony stimulating factor
coma, laboratory evaluation, 806
combination drug therapy (incl. adjunctive/add-on), 689–98
α-glucosidases, 675–6
biguanides see biguanides
dipeptidyl-peptidase-4 inhibitors see dipeptidyl-peptidase-4/DPP-4 inhibitors
GLP-1 receptor agonists see glucagon-like peptide-1 (GLP-1) receptor agonists/analouges
in hypertension, 1086
in inadquate controlled onset T2DM, 694–8
insulin see insulin administration in newly diagnosed and new onset T2DM, 689–94
sulfonylureas see sulfonylurea
thiazolidinediones see thiazolidinediones
triple, 693–4, 696–8
comorbid conditions
HbA1c tests affected by, 862–3
obstructive sleep apnea, 324–30
older adults, 839, 840
T2DM in children, 42
complications (of diabetes), 872–1101
advanced glycosylation end-products and see advanced glycosylation
bariatric surgery effects, 510–11
connective tissue, 953–63
economic cost, 1114, 1115, 1116
exercise and, 603–4
HbA1c and, 859–60
macrovascular see macrovascular disease
microvascular see microvascular complications
periodontal disease effects on, 987
retinopathy and its relationship with other types of, 892–3
stroke prevention, 1098
T1DM, acute, 24
T2DM children, 42
obstructive sleep apnea and, 327
see also comorbid conditions
computed tomography and CT angiography, arteriosclerosis, 1035
CONCEPT study, 834
confocal microscopy, corneal, 933–4
congenital malformations with diabetes in pregnancy, 827–8
detection, 831
insulin lupro and, 830
congestive heart failure, thiazolidinedione and monitoring of patients with, 668
connective tissue disorders, 953–63
constitutional factors, Rosenbloom syndrome, 960–1
contidence problems see incontinence
continuous glucose monitoring/sensing/CGM
(implantable), 769, 785, 831
added to pumps, 752
in hyperglycemic crises, 813
linked to/integrated with pumps, 750–2
pregnancy, 831
continuous infusion of glucose with model assessment (CIGMA), 341
of insulin intravenously, in hyperosmolar state and ketoadiposis management, 808–9, 813
subcutaneously see insulin pump continuous positive airway pressure (CPAP) in obstructive sleep apnea, 316, 323, 324, 327, 330, 331
contraception, postpartum, 833
contraction of muscle, 174
Convention on the Elimination of All Forms of Discrimination against Women, 1107, 1108
Convention on the Rights of the Child, 1106, 1108
Convention on the Rights of Persons with Disabilities, 1107–8
cornea confocal microscopy, 933–4
disorders, 903
coronary heart disease (CHD; CAD; coronary artery disease), 1011–15, 1064–8
atherosclerosis, 1010, 1035, 1067
clinical features, 1067–70
endothelial progenitor cells and, 1047
exercise (in diabetics) and, 603
key practical features of relationship between diabetes and, 1064–7
management aspects, 1073–4
microvascular, 1070
postprandial lipemia and, 266–7
prevalence and incidence, 1011–12
prognosis, 1065–6
revascularization after CV event, 1074
risk factors, 1013–15, 1017
see also acute coronary syndromes; myocardium, infarction
corpus cavernosum in erectile dysfunction changes within, 976–7
drugs injected into, 983–4
corticosteroids see steroids
cost (economic/financial) of diabetes care, 1113–24
direct, 1114, 1115
sulfonylureas see sulfonylurea
cystic fibrosis-related diabetes, 10, 74
DACE (Diabetes Autoimmunity Study in the Young) study, 19, 20, 21, 22, 23, 463
dapagliflozin and metformin, 693
daycare exposure to infection, T1DM and, 22
δ-deoxyglucose, 127
death, mortality, 840
dehydroyxypirandione (DHEA), 721
deafness, maternally inherited diabetes and (MIDD), 7, 427, 435, 436–7
deaths see mortalities
debriement, foot wounds, 969
decision support for older adults with diabetes, 842
DECODE (Diabetes Epidemiology: Collaborative Analysis of Diagnostic Criteria in Asia) study, 30, 31, 32, 33, 34, 35, 36, 37
DECODE (Diabetes Epidemiology: Collaborative Analysis of Diagnostic Criteria in Europe) study, 29, 30, 31, 32, 33, 35, 36, 551, 1018, 1020
dehydroyxypirandione (DHEA), 721
delay trial, 542
delivery of baby, timing and mode, 832–3
delta (b)-cells (somatostatin-producing), 61, 71, 102, 115
dementia and hypoglycemia, 630
demethylasterrginiquinone, 716
demylinating polyradiculoneuropathy, chronic inflammatory (CIDP), 931, 932
diabetes mellitus, type 1, 434
diabetic cell therapy, T1DM, 543–4
dental care
co-management of and medical care providers and providers of, 999
guidelines for practitioners, 1000
recommendations for patients, 1000
deoxyxylucose, 127
deoxyxylucose, 865
depression, 846–7
diabetes mellitus, type 1, 434
screening, 850
dermal collagen and Rosenbloom syndrome, 960
DETAIL (Diabetes Exposed to Telmsartan and Enalapril), 567
developed (higher-income) countries
drug-based therapy (pharmacotherapy) (continued)  
see also pharmacometabolomics; polypharmacy and  
specific (types of) drugs  
drug-eluting coronary stents, 1074  
drug-induced conditions  
diabetes, 8–9  
hyperosmolar state, 801, 807  
ketoacidosis, 801  
erythroid dysfunction, 979  
drug interactions  
bromocriptine-QR, 762–3  
colesevelam, 761  
thiazolidinediones, 667  
DRVS (Diabetic Retinopathy Vitrectomy Study), 902  

duodenal switch seen as biliopancreatic diversion with or without  
duodenal switch  
Dypurysyndrome (DD), 955  
limited joint mobility (Rosenbloom syndrome) vs., 957  
duration of diabetes  
and coronary heart disease risk, 1093  
and nephropathy risk, 911, 914, 917  
and retinopathy risk, 892  
dysglycemia and obstructive sleep apnea, 316,  
319–24, 331  
see also hyperglycemia; hypoglycemia  
dyslipidemia (hyperlipidaemia/fatty acid metabolic  
disturbances), 262–74, 880–1  
bariatric surgery effects, 511  
cardiovascular disease and  
atherosclerosis, 1032, 1038–40  
coronary heart disease risk, 1014  
colesevelam, 759  
HbA1c assays and, 863  
inulin gene expression and, 86–7  
in nephropathy prevention/management, 568,  
921–2  
obesity and, 880–1  
in retinopathy prevention, management, 570  
of T2DM and insulin resistance, 262–74, 278, 342  
therapy, 265–70, 722  
see also hypertriglyceridaemia  
dysmotility see motility disorders  
dysrhythmias see arrhythmias  
E  
early diagnosis and detection, 1120  
nonalcoholic fatty liver disease, 292, 301  
peripheral vascular disease, 1091, 1092  
T2DM, 1119, 1121  
eating disorders, 848  
screening, 850  
echocardiography, heart failure, 1070  
edema  
cerebral, complicating diabetic ketoacidosis, 812  
macular see macular edema  
EDIC (Epidemiology of Diabetes Interventions and  
Complications study), 565, 566, 568, 767, 859,  
876, 920, 940, 948, 1034  
education and information (patient), 1108–10  
foot ulcers, 971  
hypoglycemia prevention, 793  
older adults, 841–2  
rights and responsibilities on, 1108–10  
eggs, optimal amounts, 585  
EF2AK3 genes and Wolcott- Rallison syndrome, 427,  
434  
elderly see aging persons  
electrolyte metabolism and  
ketoacidosis and  
hyperosmolar state, 803  
electron transport chain see respiration, mitochondrial  
electrophoretic methods  
HbA1c measurement, 861  
nephropathy detection, 913–14  
ELMO 1 gene, 913  
embryology, β-cell, 56–9  
embryosma, intestinal, 680  
empowerment, 1105, 1106, 1108, 1109, 1111  
pyramid of, 1110  
endoplasmic reticulum (ER)  
gluco- and, 86  
monogenic disorders of β-cell, 427  
proinsulin in biosynthesis, 87  
transport to Golgi apparatus from, 89  
stress, 86, 882, 1040  
endoscopic techniques of bariatric surgery, 506  
endothelium, 1046–50  
cardiovascular disease and, 1047  
atherosclerosis, 1035–40, 1046  
erythrocytes (red blood cells) and glucosetransport,  
474, 476, 521  
endovascular (intravascular) ultrasound,  
atherosclerosis, 1035  
energy (fuel), 479–88, 581–3  
balance and homeostasis, 163, 311, 509, 522, 533,  
534  
brain see brain  
energy metabolism and its role in regulation of,  
479–88  
glucokinase in, 205  
obesity and T2DM, 473  
dietary intake see caloric content and intake  
distribution from macronutrients, 581–3  
expedurance  
analysis of data in humans, 485  
bariatric surgery effects, 509, 512–13  
components and their relevance to obesity, 481–2  
drugs targeting, 484–5  
in etiology of obesity, 482–3  
inulin effects, 531–2  
leptin effects, 531–2  
methods of measuring, 479–81  
hepatic energetics, 202–3  
enaosarin thromboprophylaxis, 1057  
tenteropathy, 946  
tenteroviruses, 22, 446–7, 461  
environmental factors  
elderly persons, 837  
in metabolic syndrome, gene interactions with, 1033  
in obesity, animal studies, 520–1  
in T1DM, 22  
interaction with, 23, 446–7  
in T2DM  
animal studies, 520–1  
genetic interactions with, 408–9  
see also lifestyle  
Epc2 (2A), 101, 618, 635  
EPIC see European Prospective Investigation into  
Cancer  
epidemic of diabetes, 401–2, 1105  
epidemiology, 17–51, 989  
autonomic neuropathy, 939  
cardiovascular, 941  
gastrointestinal, 944  
cerebrovascular disease and stroke, 1015–16, 1097  
coronary heart disease, 1011–12  
erectile dysfunction, 975–6  
foot ulcers, 964–5  
gestational diabetes, 824  
hyperosmolar state, 805  
hypertension, 1007–8  
ketoacidosis, 799, 800  
low-income countries, 1122  
obstructive sleep apnea, 317–18  
older adults with diabetes, 836–7  
periodontal disease, 988, 989  
peripheral vascular disease, 1091–2  
retinopathy, 890–1  
T1DM, 17–28  
T2DM, 29–51  
atelencephaly, 817  
cancer and, 306–8  
see also burden; incidence; prevalence  
epigenetics, 876–7  
atherogenesis, 1038  
nephropathy, 913  
obesity, 520  
T2DM, 409, 520  
therapeutic targeting, 877  
epinephrine and insulin gene expression, 85, 88  
epithalamus (gene–gene interactions) in  
T2DM-associated, 408  
epithelium in pancreatic development, 58–9  
erectile dysfunction, 946–8, 965–87  
assessment, 943, 980  
epidemiology, 975–6  
management, 947, 980–5  
ERKs (extracellular-regulated kinases), 83, 84, 179,  
313, 382, 883  
erythrocytes (red blood cells) and glucose transport, 211  
erythrophoetin inhibition in retinopathy, 893  
endophagal dysmotility, 945  
estrogen and cancer and T2DM, 310  
eteinercept, 543  
ETDRS (Early Treatment Diabetic Retinopathy  
Study), 899  
classification of retinopathy, 894, 896  
etnicity and race  
glucose dysregulation, 35–6  
nonalcoholic fatty liver disease, 287  
obstructive sleep apnea, 317  
retinopathy and, 892  
T1DM, 18  
T2DM, 29–35  
etiology factors see causes  
EUCLID (EURODIAB Controlled Trial of Lisinopril  
in Insulin-Dependent Diabetes Mellitus), 898  
euglycemic insulin (hyperinsulinemic–euglycemic)  
clamp technique, 195, 221, 235–6, 239, 242, 340,  
341  
EUROBIRD ("European Best Information through  
Regional Outcomes in Diabetes"), 1111  
EURODIAB studies, 18, 800, 967, 1008  
EURODIAB Controlled Trial of Lisinopril in  
Insulin-Dependent Diabetes Mellitus (EUCLID), 898  
Europe  
dietary management in diabetes, 577–85
fuel see energy
Funagata study, 1018
functional food, 578, 583
functional MRI (of brain) with ingested glucose load, 388 in peripheral neuropathy, 965
G
g-cells, gastrin production from, 153
G-protein-coupled receptors (GPCRs; GPRs), 101–2, 524–5
fatty acid binding to, 524–5
drugs acting as agonists at, 715–16
Ga/ε-protein, 102
GPR40, 102, 176, 524, 715
GPR91, 880
GPR119, 102, 148, 524, 715, 716
GPR120, 524
Gq-protein, 102
Gs-protein, 101–2
Gα1, 168
galanin, 129–30
Galega officinalis and galegine, 641
gallstones, cholesterol, in nonalcoholic fatty liver disease, 289
gamma-glutamyltransferase (GGT) and insulin resistance, 343, 344
gas exchange, measurement (direct calorimetry), 479–80
gastrectomy, vertical banded (VSG), 505–6, 507
glycemic control following, 510
insulin secretion and β-cell function after, 512
mechanism of weight loss, 586
microvascular complications after, 511
remission, 510
gastric banding, laparoscopic adjustable (AGB/LAGB), 305–6
glycemic control, 510
mechanisms of weight loss, 508
microvascular complications after, 511
nonalcoholic fatty liver disease, 296, 297
remission, 510
gastric bypass see Roux-en-Y gastric bypass
gastric function (incl. emptying)
autonomic neuropathy and, 942, 944, 945, 946
GLP-1 effects, 150–1, 727, 740
GLP-1 receptor agonists effects, 731, 946
gastric inhibitory peptide see glucose-dependent insulinotropic polypeptide
gastrin, 153
gastrin-releasing polypeptide (GRP), 126–7
gastrointestinal tract see gut
gastraparesis, 945–6
gastroplasty, vertical banded (VBY), 505–6, 508, 509
gastrulation, 56, 67
gemfibrozil, 266, 270, 1096
gender see men; sex; women
genes see environmental factors
environmental factors interacting with see environmental factors
expression in glucagon regulation, 198–9
insulin regulation of, 182–4
hyperglycemia and regulation of, 876–7
in pancreatic growth and development, 59–60, 61
polymorphisms see polymorphisms
T2DM-associated, 402–9
candidate genes, 403–4
functions, 404–7
high-density mapping, 404
see also association studies; linkage studies; transcription and specific genes
gene therapy, erectile dysfunction, 985
generalized anxiety disorder (GAD), 847
genetic defects and factors
aging and, 838
causing diabetes, 7–10, 426–41
environment—gene interactions, 23–4, 1033
immune-mediated diabetes, 443–6
insulin secretion impairment, 375–6
monogenic see monogenic diabetes
polygenic see polygenic disorder
T1DM, 20, 443–6, 454–6
T2DM see type 2 DM
types/classification of disorders, 6, 7–8, 10–11
see also monogenic diabetes; polygenic disorder
causing hypertension, 1080
causing nephropathy, 913, 917
causing nonalcoholic fatty liver disease, 287–8
causing retinopathy, 892
see also epigenetics; molecular genetics and specific genes
genetic models of obesity, 521–2
cancer in, 311–12
of T2DM, 521–2
genetic testing for neonatal diabetes, 435–6
genitourinary autonomic neuropathy, 943, 946–8
genome-wide association studies
T1DM, 460
T2DM, 404
genotype, thrifty, 402, 551
geographic distribution
glucose dysregulation, 35–6
retinopathy, 890–1
T1DM, 18
T2DM, age and sex-specific, 30–5
see also developed countries; developing countries
geriatric syndromes, 839–41
German SWEETHEART Registry, 1065, 1066
gestational age
determination, 831
large-for-, 828
gestational diabetes, 823–4
classification, 5
diagnosis, 13, 824–6
glibenclamide compared with or as alternative to insulin in, 626
screening, 824–5, 858
glibenclamide, 102, 472
cells producing, 71
gingiva (gums)
health check at physician's office, 1000
inflammation (gingivitis), 992
glucoma, 903
glia–axon dysjunction, 929
glibenclamide, 626
hypoglycemia with, 626, 629
neonatal diabetes, 434
glucagon, 626–7
modified-release, 627
glimepiride, 627–8
glinides see meglitinides
glimepiride, 627–8
compared with glibenclamide, 626
glutathione peroxidase-4 (GPx4) inhibitors
glipizide (PPARγ agonists), 270, 657, 660, 719–20
Global Burden of Diseases, Injuries, and Risk Factors Study (GBD), 988
global perspectives of economics of diabetes care, 1113–24
glomerulus filtration rate (GFR), determination, 919
Index

glucose-lowering (hypoglycemic) agents (continued)
retinopathy, 898
in pregnancy/see pregnancy
subcutaneous see subcutaneous administration
thrombosis prevention, 1058, 1059
see also specific (types of) glucose-lowering drugs
glucose tolerance
impaired see glucose intolerance
unimpaired, HbA1c in patients with, 855–7
glucose tolerance test
intravenous see intravenous glucose tolerance test
oral see oral glucose tolerance test
glucose transporters (GLUTs/SLC2A), 175–9, 211, 226
GLUT1, 96, 211, 212, 383
biguanides and, 644
GLUT2 (SLC2A2A), 96, 175, 204, 212, 226, 383
mutations, 426, 433
GLUT4, 163, 169, 172–81, 211, 213, 226–81, 383
Akt and, 171, 172, 174
biguanides and, 644
cardiomyopathy and, 1069
protein metabolism and, 252–3
storage vesicles (GSVs), 176, 177, 178, 179, 213
training in healthy subjects and, 598
training in T2DM and, 601
vesicle translocation and trafficking, 175–8
glucose-6-phosphate (G6P), 195, 196, 198, 203, 212
glucose-6-phosphatase (G6Pase), 183, 198, 199, 205,
glucohexokinase (HK), 359
GLUT1, 96, 194, 195, 198, 211, 212
GLUT4, 177, 213
training in T2DM and, 600, 601
training in healthy subjects and, 598
storage vesicles (GSVs), 176, 177, 178, 179, 213
protein metabolism and, 252–3
Akt and, 171, 172, 174
GSK-3, 87, 171, 196, 205, 206, 225, 381, 722
inhibitors, 722
bedside point-of-care monitoring with hyperglycemic crises, 813
classification related to levels of, 4
continuous monitoring see continuous glucose monitoring
treatment (interventional) see glycemic control treatment (physiologic) see glucose, homeostasis conversion factors for glucose concentrations, 29–30
estimated average glucose (eAG) in HbA1c assessment, 856–7
vs. mean blood glucose, 857
fasting see fasting glucose
self-monitoring (SMBG), 749–50, 751, 752, 856
HbA1c as adjunct to, in pregnancy, 859, 864–5
see also dyshyperglycemia; hyperglycemia; hypoglycemia;
normoglycemia
glycemic control (interventional)
α-glucosidase inhibitors and, 674, 675
biguanides and, 645
cardiovascular disease and, 1033–4
collesevalam and, 758–9
in diabetic ketoacidosis, targets, 810, 813
erectile dysfunction and, 980–1
exercise and, 603, 606
gastric emptying in improvement of, modulation of, 946
HbA1c in 1DM as unique index of, 858
intensive, recent emphasis on, 630
metabolic memory phenomenon see metabolic memory
microvascular disease prevention and, 375
neuropathy prevention, 565
neuropathy prevention, 568–70
elderly adults, 839
periodontal care and, 998
periodontitis and poor levels of, 993–4
in pregnancy, 828–9
adverse outcomes related to poor levels of, 837–8
in labor, 833
retinopathy prevention, 570
sulfonlureas and, 624, 625, 630
tiazolidinediones and, 660
glycemic index, 44
current recommendations in diabetes, 578, 580–1
means for various foods, 582
glycerol and gluconeogenesis, 197
glycogen breakdown (glycogenolysis) in liver, 194, 195–6
drugs affecting, 722
exercise and, 590–1
ketoacidosis and, 802
regulation, 196
synthesis, 227, 381, 384–5
ketoacidosis and, 802
liver, 195, 196, 205–6
training and, 598
therapeutic targeting of metabolism, 206
glycogen phosphorylase, 196, 206, 225, 381
exercise and, 591–2
inhibitors, 722
glycogen synthase (GS) 3, 87, 171, 196, 205, 206, 225, 381, 722
inhibitors, 722
glycogen synthase kinase (GSK)
GSK-3, 87, 171, 196, 205, 206, 225, 381, 722
inhibitors, 722
gastrointestinal tract (includ. stomach and intestine)
biguanides and, 644
cardiomyopathy and, 1069
cardiomyopathy and, 1069
protein metabolism and, 252–3
storage vesicles (GSVs), 176, 177, 178, 179, 213
training in healthy subjects and, 598
training in T2DM and, 601
vesicle translocation and trafficking, 175–8
glucose-6-phosphate (G6Pase), 183, 198, 199, 205,
212, 644
drugs affecting/inhibiting, 644, 722
glucose-6-phosphate (G6P), 195, 196, 198, 203, 205
future cycling, 219
α-glucosidase (AGIs) and α-amylase inhibitors, 673–85
clinical applications, 674–8
prevention of T2DM, 556, 676–7
weight loss properties, 500, 678
combined with other drugs, 675–6
contraindications, 680
dose recommendations, 681
dose–response relationships, 675
existing approved, 712
indications, 680–1
mechanism of action, 673–4
other beneficial effects, 678–9
reversible competitive, 710
tolerability/ adverse effects, 679–80
see also specific drugs
glucotoxicity (of chronic hyperglycemia), 85–6, 87,
376, 413–25, 1036–8
β-cell and
failure, 420
loss, 75
clinical significance, 421
definition and terminology, 413
molecular mechanisms, 417–21
GLUT see glucose transporters
 glutamatic acid decarboxylase (GAD) antigen, 4, 20,
401, 543
vaccine, 543
glyburide
and metformin in T2DM, 646, 690
in pregnancy, 830
glycation, 854
advanced see advanced glycosylation hardening, 854
hemoglobin A1c (HbA1c) see hemoglobin A1c
reactions, 854
serum proteins, 863–4
thrombosis and effects of, 1056
glycemia (blood/plasma glucose)
age- and sex–specific concentrations, 30
1,5-anhydroglucitol as measure of, 865
H
hand(s), disorders affecting, 955–7, 958
see also fingers
Handan Eye study, 891
HAPO (Hyperglycemia and Adverse Pregnancy Outcome) study, 13, 826, 858
haplotyping and hypertension, 1080
HDL see high-density lipoprotein
health policy-makers, cost-effectiveness and, 1119
right to, 1106–7, 1107, 1108, 1110
health insurance, USA, 1116–17
hyperglycemic hyperosmolar state (HHS) (continued)
epidemiology, 805
mortality, 807–13
future prospects, 813
hyperglycemic crises see hyperglycemia
mortalities, 799, 805
pathogenesis, 807
precipitating and predisposing factors, 801, 805–6
water and electrolyte metabolism, 803
hyperinsulinemia, 234, 1009
hypoglycemia, 783–94
in nonalcoholic fatty liver disease, 288–9
thrombosis and, 1056
hypocaloric diet in nonalcoholic fatty liver disease, 289
hypoglycemia, 783–94
acute episodes, 789, 791–3
asymptomatic, 784, 792
cardiovascular risk (in severe/critical hypoglycemia), 790, 1070, 1071
arrhythmias, 790, 1070, 1071
coma, laboratory values, 806
consequences (of severe hypoglycemia), 629–30, 789–90, 1070–3
cardiovascular see subheading above
definitions, 783–4
exercise-related, in T1DM and T2DM, 602
impaired awareness, 784, 786–7, 788, 790, 791
barriers to restoration of awareness, 794
restoration of awareness, 793–4
insulin-induced, 698, 783–94
assessing and managing, 791–4
barriers to avoidance of, 794
complications, 783–98
in diabetic ketoacidosis treatment, 812
with insulin pump, 760, 767–8, 768, 769
insulin pumps in reduction of, 748, 749, 751, 753
prevention of future events, 793–4
moderate, 784
nocturnal see nocturnal hypoglycemia
oral glucose-lowering drug-induced α-glucosidase inhibitors, 679–80
cardiovascular event risk, 1073
macrovacular disease/complications (continued)
biguanides, 643 – 8
see also cardiovascular disease and specific types of macrovacular complications
macular edema, 894, 897
treatment, 899, 906, 901, 902
MAA, 61, 83 – 4, 84, 85, 86, 87, 420
MAL, 61, 83 – 4
MAGIC (Meta-Analyses of Glucose- and Insulin-related traits Consortium), 404, 407
magnesium
hypertension and, 1081
supplements, 720
magnetic resonance imaging
Charcot neuroarthropathy, 972
functional see functional MRI
Maillard reaction, 854, 882, 960
Malta, 61, 83 – 4
MALDI-MSI (Matrix-Assisted Laser Desorption/Ionization and Mass Spectrometry Imaging), 709
malignant lymphoma
centrosarcoma (Kaposi sarcoma), 587
retinopathy, 893, 898 – 902
bone, 1050 – 51
osteosarcoma, 1050
multifactorial
T1DM and T2DM see type 1 DM; type 2 DM
mandibular advancement devices, 330
MAP kinase see mitogen activated protein kinase
Mapuche of Chile, 35
mass spectrometry and metabolomics, 277
MATCH (Management of Atherothrombosis with Clopidogrel in High-risk patients with recent TIA or ischemic stroke) Trial, 1098
maternal…[as in pregnancy] see pregnancy
maternally inherited diabetes and deafness (MIDD), 7, 427, 435, 436 – 7
maturity-onset diabetes of the young (MODY), 374, 427 – 33
diagnosis, 428
mutations causing, 7
glucokinase (GCK-MODY), 7, 374, 426, 428 – 30
transcription factor (TF-MODY), 374, 426, 428, 429, 431 – 3
subtypes/phenoatypes, 428 – 33
MCP-1 (monocyte chemoattractant protein), 1041
meat, optimal amounts, 585
median nerve compression at carpal tunnel (= carpal tunnel syndrome), 955, 958
medical care (and periodontal disease) costs, 987 – 8
dental care and, co-management between providers of, 999
Medical Research Council/British Heart Foundation Heart Protection Study (MRC/BHF HPS), 267
Medtronic Minimized implantable insulin pump, 753, 754 – 5, 769
medullary thyroid carcinoma and GPL-1 agonists, 738
meglitinides (glinides), 615 – 40, 711
chemistry and structure, 618 – 19
combined with other metformin, 689, 690 – 1, 694 – 5
existing approved, 712
mechanism of action, 616 – 18
prevention of T2DM, 556
in “treat-to-tail” approach, 694 – 5
Malaria, 61, 83 – 4
Malmö trial in Sweden, 43, 375, 375
malnutrition
adult, diabetes related to, 5
maternal and fetal, 377, 520
maternal and infant, 5
Malmquist, 357, 875 – 871
albunin levels in urine see albuminuria; macroalbuminuria; microalbuminuria
autonomic neuropathy
anoxic neuropathy, 945 – 6
cardiovascular, 943 – 4
cellular risk related to, 1070 – 3
cell-based see cell-based therapies
Charcot neuroarthropathy, 972 – 3
collaborative see multidisciplinary team
coronary heart disease, 1073 – 4
cost see cost; cost-effectiveness; cost-utility
dyslipidemia, 265 – 70, 722
erectile dysfunction, 947, 980 – 5
foot ulcers, 969 – 71
hyperglycemic hyperosmolar state see hyperglycemic hyperosmolar state
hypertension see hypertension
hypoglycemia, 791 – 4
ketoadiposis see ketoacidosis
less-than-optimal, consequences, 1117
lifelong, 1001
lifestyle see lifestyle
maternally inherited diabetes and deafness, 437
maturity-onset diabetes of the young
glucokinase MODY, 430
transcription factor MODY, 432
metabolic syndrome, 350
neonatal diabetes
permanent, 434
transient, 435
nephropathy management, 919 – 23
nonalcoholic fatty liver disease, 292 – 305
obesity, see also obesity
obstructive sleep apnea, 330
older adults see aging persons
patient-led see self-management
periodontal disease, effects on diabetes of, 996 – 7
peripheral vascular disease, 1093 – 7
pharmacological see drug-based prevention;
drug-based therapy
in pregnancy see pregnancy
quality, improving, 1117, 1118
renal cysts and diabetes syndrome, 436
retinopathy, 897, 898 – 902
surgical, see also surgery
T1DM and T2DM see type 1 DM; type 2 DM
mandibular advancement devices, 330
MAP kinase see mitogen activated protein kinase
Mapuche of Chile, 35
mass spectrometry and metabolomics, 277
MATCH (Management of Atherothrombosis with Clopidogrel in High-risk patients with recent TIA or ischemic stroke) Trial, 1098
maternal…[as in pregnancy] see pregnancy
maternally inherited diabetes and deafness (MIDD), 7, 427, 435, 436 – 7
maturity-onset diabetes of the young (MODY), 374, 427 – 33
diagnosis, 428
mutations causing, 7
glucokinase (GCK-MODY), 7, 374, 426, 428 – 30
transcription factor (TF-MODY), 374, 426, 428, 429, 431 – 3
subtypes/phenoatypes, 428 – 33
MCP-1 (monocyte chemoattractant protein), 1041
meat, optimal amounts, 585
median nerve compression at carpal tunnel (= carpal tunnel syndrome), 955, 958
medical care (and periodontal disease) costs, 987 – 8
dental care and, co-management between providers of, 999
Medical Research Council/British Heart Foundation Heart Protection Study (MRC/BHF HPS), 267
Medtronic Minimized implantable insulin pump, 753, 754 – 5, 769
medullary thyroid carcinoma and GPL-1 agonists, 738
meglitinides (glinides), 615 – 40, 711
chemistry and structure, 618 – 19
combined with other metformin, 689, 690 – 1, 694 – 5
existing approved, 712
mechanism of action, 616 – 18
prevention of T2DM, 556
in “treat-to-tail” approach, 694 – 5
Melanin, T1DM age and sex-specific prevalence, 34
melatonin receptors, 407, 1080
Melbourne Visual Impairment Project, 890
men
coronary heart disease, 1012
acute myocardial infarction, 1013
ketoacidosis and, 801 – 3
thiazolidinediones, 662 – 4
in exercise, 590 – 2
in diabetics, 592 – 6
protein, 256 – 7
glycagone, 139, 217
ketoadiposis and, 801 – 3
older adults, alterations, 837
peroxisome proliferator-activated receptor effects, 658
PI-3 kinase and insulin effects on, 169 – 71
retinopathy and, 893
Rosenbloom syndrome and, 960 – 1
see also specific substrates
metabolomics and T2DM and insulin resistance, 975 – 80
metacarpophalangeal joints (MCPs) in Rosenbloom syndrome, 956, 957
metformin see biguanides
methylation
DNA, 409, 435, 876, 913
histone, 520, 876
methylglyoxal, 883
microalbuminuria, 346, 911, 914, 915, 916, 1008, 1085
exercise and, 604
management, 567, 585, 665, 918 – 19, 920, 921, 1085
obstructive sleep apnea and, 329
screening for, 918 – 19
microbiota, 473 – 6
bariatric surgery effects, 513
in health, 473
obesity and T2DM and, 473 – 6
animal studies, 520 – 1
and dietary interventions, 475 – 6
short chain fatty acids and, 474, 524 – 5
Micronesia, T2DM age and sex-specific prevalence, 34
micronutrients, dietetic recommendations, 583
supplements, 720
see also minerals; vitamins
microparticles, endobiotic release, 1047
microRNAs, 408, 877
nephropathy and, 917
microsomal triglyceride transfer protein (MTP), 263, 264
microvascular complications/disease, 875 – 951
bariatric surgery effects, 510 – 11
HbA1c and, 859
heart, 1070
neuropathy genesis and the role of, 930
obstructive sleep apnea and, 327
pathogenesis see pathogenesis
prevention/risk reduction, 564 – 73
α-glucosidase inhibitors, 678 – 9
biguanides, 646 – 8, 877
strict glycemic control in see glycemic control
Rosenbloom syndrome associated with, 959 – 60
see also specific microvascular complications
Middle East, T2DM age and sex-specific prevalence, 34
miglitol, 673, 674, 675, 760, 710
nervous system

autoimmune neuropathy see autonomous nerves

β-cell response influenced by, 115–16

central see brain (incl. cerebrum and central nervous system)

safety/ appetite and, 508–9

see also neuroendocrine system

neuritis, painful peripheral, drugs causing erectile dysfunction in, 979

neuroarthopathy, Charcot, 972–3

neurocognitive impairment see cognitive impairment

Neurol (Beta2), 61, 83, 84, 85

neuroendocrine system in exercise, 591

in polycystic ovary syndrome, defects, 591

Neurop3, 60–1

neuroglycopenia, 785–6

neurologie assessment of diabetic foot, 968

neuropathy (autonomic) see autonomic neuropathy

neuropathy (peripheral and in general or unspecified), 568–70, 926–37, 965–6

neuralgia, 928–9

nerve fibers (axons) in neuropathy

atrophy, 928–9

glia dysfunction, 929

intradermal (IENF), 932–3, 934

myelinated, 926–8

unmyelinated, 930

neuritis, painful peripheral, drugs causing erectile dysfunction in, 979

neuroarthopathy, Charcot, 972–3

neurocognitive impairment see cognitive impairment

Neurol (Beta2), 61, 83, 84, 85

neuroendocrine system in exercise, 591

in polycystic ovary syndrome, defects, 591

Neurop3, 60–1

neuroglycopenia, 785–6

neurologie assessment of diabetic foot, 968

neuropathy (autonomic) see autonomic neuropathy

neuropathy (peripheral and in general or unspecified), 568–70, 926–37, 965–6

neuralgia, 928–9

nerve fibers (axons) in neuropathy

atrophy, 928–9

glia dysfunction, 929

intradermal (IENF), 932–3, 934

myelinated, 926–8

unmyelinated, 930

PARACRINOPATHY OF ISLETS OF LANGERHANS, DIABETES, AS, 136, 141
PARACELLULAR (C-CELL) CARCINOMA AND GLP-1 AGONISTS, 738
PARASYPATHETIC NERVES ERECTILE DYSFUNCTION AND, 976
ISLETS, 125–6, 128
PHYSiologic ROLES, 132
SEE ALSO AUTONOMIC NERVES; PARASYMPATHETIC NERVES; SYMPATHETIC NERVES
PARICALCITOL IN NEPHROPATHY, 923
PARIS PROSPECTIVE STUDY, 1012, 1014, 1018
PAS (PER-ARNT-SIM) KINASE, 84, 87
PASTA, GLYcemic INDEX, 582
PATATIN-LIKE PHOSPHOLIPASE DOMAIN CONTAINING 3 PROTEIN (PNPLA3) II48M VARIANT, 286, 287–8
PATHOGENESIS (AND PATHOPHYSIOLOGY), 339–46, 466–7
1036–42
ATHEROSCLEROSIS, 1035–42
AUTONOMIC NEUROPATHY, 939–41
CARDIAC, 1067
GASTROINTESTINAL, 944
HYPERTENSION IN DIABETICS, 1008–9
IN ELDERLY PERSONS, 837
FOOT ULCERS, 964–6
HYPERGLYCEMIA SEE HYPERGLYCEMIA
HYPEROSMOLAR STATE, 811
HYPEROSMOLAR HYPERGLYCEMIC KETOACIDOSIS, 801
METABOLIC SYNDROME, 1033, 1036–8
MICROVASCULAR COMPLICATIONS, 375–88
RETINOPATHY, 893
NONALCOHOLIC FATTY LIVER DISEASE, 284–8
OBSTRUCTIVE SLEEP APNEA (OSA), 318
PERIPHERAL NERVES, 28–1
PERIPHERAL IMMUNETOLERANCE, 446
PERIPHERAL NEUROPATHY, 926, 931, 965
T1DM, 422–66
T2DM, 339–42, 687–8
IMPLICATIONS FOR TREATMENT, 388, 688–9
OBESITY AND CHILDHOOD/ADOLESCENCE AND, 819
OBESITY AND GUARDIAN FACTORS AND, 469–78
THROMBOSIS SEE THROMBOSIS
SEE ALSO CAUSES
PC(1) SEE PROHORMONE CONVERTASES
PC(1) PLASMA DIFFERENTIATION FACTOR–1, 182
PDGFB OINTMENT IN FOOT ULCER THERAPY, 970
PEDRO-1 (PANCREATIC DUODENAL HOMEBOX-1), 58, 61, 64, 83, 84, 85, 86, 87, 714
NEONATAL DIABETES AND, 433
PEDIATRIC DISEASE SEE CHILDREN PEDOMETERS, 480
PEGS, 718
PERIPHERAL NERVE INJURY, 738
PERIPHERAL NERVES SEE ERECTILE DYSFUNCTION
PEROXISOME PROLIFERATOR-ACTIVATED RECEPTOR-γ (PPARγ), 658–60, 719–20
AGONISTS SEE THIAZOLIDINE NIDINES CO-ACTIVATOR 1 (PPAR-γ), 183, 198, 199–200, 200
SEE ALSO THIAZOLIDINE NIDINES CO-ACTIVATOR 1 (PPAR-γ), 183, 198, 199–200, 200
PERITONEAL DIALYSIS AND EXERCISE, 604
PEROXISOME PROLIFERATOR-ACTIVATED RECEPTOR-α (PPARα), 658, 1069
AGING SEE FIBRINOLYTIC AND, 1069
OMEGA-3 POLYUNSATURATED FATTY ACID EFFECTS, 298
PEROXISOME PROLIFERATOR-ACTIVATED RECEPTOR-α/γ (PPARα/PPARγ) AGONISTS, 270, 657, 660, 719–20
PEROXISOME PROLIFERATOR-ACTIVATED RECEPTOR-γ (PPARγ), 658–60, 719–20
AGONISTS SEE THIAZOLIDINEDIONES CO-ACTIVATOR 1 (PPAR-γ), 183, 198, 199–200, 200
PGC-1 (PPAR-α CO-ACTIVATOR, 1) 183, 198, 199–200, 200
PHENOLIC COMPOUNDS, 279
PHOSPHATIDYLINOSITOL 3-KINASE (PI3K), 102–3, 169–72
PHOSPHATIDYLINOSITOL 4,5,6-TRIPHOSPHATE (PIP3), 530, 658
PHOSPHATIDYLCHOLINE, 1091–7
PHOSPHOLIPID 2002–6, 727, 385
PHOSPHOFRUCTOKINASE (PFK), 202, 204–6, 727, 385
PHOSPHOFRUCTOKINASE (PFK), 202, 204–6, 727, 385
PHOSPHOFRUCTOKINASE (PFK), 202, 204–6, 727, 385
PHOSPHOFRUCTOKINASE (PFK), 202, 204–6, 727, 385
lifestyle factors in, 43, 267, 495 – 6, 551 – 4, 557, 838
primary care and, 558
program implementation, 559
see also primary prevention
primary care and T2DM prevention, 558
primary prevention
cost-saving and cost-effective interventions for, 1119 – 21
T1DM, 544 – 5
T2DM, 1121
Princeton consensus guidelines (2nd) on cardiovascular risk stratification in patients with sexual dysfunction, 981, 982
PRxActive Study (PROspective pioglitAzone Clinical Trial In macroVascular Events), 69, 664, 665, 666, 669, 696, 699
progenitor cells
endothelial, 1047 – 8
pancreatic/enocrine, 59 – 62
multipotent (MPCs), 59, 60, 64
see also stem cell therapy
prognosis of coronary heart disease in diabetes, 1065 – 6
PROGRESS (Perindopril Protection Against Recurrent Stroke Study), 1097
progression see natural history
prohormone convertases (PCs)
proglucagon, 138
proinsulin
biosynthesis, 87 – 9
processing/conversion to insulin, 89 – 91
dysfunctional, 92
structure, 87
sulfonylurea effects on secretion and synthesis, 622
transport from ER to Golgi apparatus, 89
proinsulin conversion to insulin, 89 – 91
prokaryotic, 784
proliferative retinopathy, 881–2
psychological problems, 846 – 52
elderly, 840
erectile dysfunction caused by, 979
psychosocial counselling, 984 – 5
PTEN22 gene, 20, 445, 446, 448, 460, 461, 462
puberty
retinopathy onset, 892
T2DM onset, 41
see also adolescents
pulmonary... see lung
pulse (arterial), palpation in peripheral vascular disease, 1092
pulses (legume), optimal amounts, 584
pumps see insulin pumps
pupillary abnormalities, 948
pumps see insulin pumps
purinergic P2Y receptor inhibitors in
thromboprophylaxis, 1058 – 9
pyruvate dehydrogenase (PDH), 200 – 1, 207 – 8, 227, 385, 1069
pyruvate kinase, 201 – 2, 207
Q
Qsymia® see phentermine/topiramate
QT prolongation, 631, 941, 943, 946, 1070, 1071
quality of care, improving, 1117, 1118
quality of life (QoL) outcomes
acarbose and, 679
insulin pumps, 753
children, 747
elderly, 748
pancreatectomy and islet transplantation, 780
quantitative insulin sensitivity check index (QUICKI), 241, 342
quantitative sdomotor axon reflex test, 948
Quantocep™, 279
QUICKI, 241, 342
R
Rab GTPases, 100, 174, 177 – 8
race see ethnicity and race
radiculoplexus neuropathy, diabetic lumbosacral (DLPN), 931, 932
radiographs see X-ray plain films
radionuclide imaging (scintigraphy), coronary heart disease, 1068
RAGE (receptors for AGEs), 886, 893, 913, 941, 1037
ramipril
hypertension, 1083
nephropathy prevention, 567
T2DM prevention, 556
Randle (cycle) hypothesis, 224, 385, 722, 1069
ranitizumbib, 571, 900 – 1, 901
ocular and systemic safety, 901
rapamycin, 543
Ras, 165, 168
RASS (Renin – Angiotensin System Study), 898, 921
reactive species of oxygen see free radicals
Reaven’s syndrome see metabolic syndrome
RECORD (Rosiglitazone Evaluated for Cardiovascular Outcomes and Regulation of Glycemia in Diabetes) study, 657
red blood cells and glucose transport, 211
refractive error, 902
regeneration, β-cell, 73, 356 – 8
regulatory T cells in T1DM therapy, 544
Relationship between Insulin Sensitivity and Cardiovascular disease (RISC) study, 346
relatives see family (and relatives)
relaxation training, 847
remission
after bariatric surgery, 509 – 10
mechanisms, 511 – 13
after clinical onset see honeymoon period
RENAAL (Reduction of Endpoints in NIDDM with the Angiotensin II Antagonist Losartan), 566, 567, 921
renal... see kidney; nephropathy
renin, 880
inhibitors (direct) in nephropathy prevention, 567, 880
renin – angiotensin – aldosterone (RAAS) system
coronary heart disease and, 1015
microvascular complications (and their prevention) and, 879 – 80
nephropathy, 566 – 8, 915, 918, 921, 923
retinopathy, 893, 898
see also aldosterone; angiotensin; angiotensin-converting enzyme
renin – angiotensin – aldosterone (RAAS) system
blocks/ inhibitors
hypertension and, 1085, 1086
kidney disease and, 918, 921, 923, 1085 – 6
T2DM prevention, 556
Renin – Angiotensin System Study (RASS), 898, 921
repaglinide, 628
repetinorm and, 689, 691
structure, 617
replication, β-cell, 64, 65
resistin, 719, 1041
respiration, mitochondrial (incl. electron transport chain)
β-cells, 97, 620, 643, 878
bipigandus and, 643 – 4
respiratory quotient, 219, 224, 484
resting metabolic rate, 481
RESTORE trial, 900, 901
retinal vessels
occlusion (arteries and veins), 903
retinopathy-related changes, 903
retinoblastoma protein 4 (RBP4), 719
retinopathy (DR), 889 – 92
clinical features and assessment, 894
epidemiology, 890 – 1
exercise and, 694
HBAlc and, 859, 898
investigations, 895 – 6
management/treatment, 897, 898 – 902
natural history (compared with T2DM), 914–15
nonalcoholic fatty liver disease and risk of, 289
obesity-linked see obesity-linked type 2 diabetes
obstructive sleep apnea and, 324–9
impact of T2DM, 331
osteonpenia, 954
pathogenesis see pathogenesis
postprandial glucose disposal in, consequences of acute stimulatory effect of hyperglycemia on, 414
prevention see prevention
proinsulin processing dysfunction, 92
protein metabolism in, 257–8
psychological problems
anxiety, 848
depression, 846
eating disorders, 848, 849
neurocognitive problems, 849
screening, 1121
treatment/management
drugs see subheading above
obesity-driven T2DM in youth, 820–1
pathophysiology and its implications for, 388, 688–9
periodontitis and effects of, 995
physical activity, 496
weight loss in, 486
tyrosine kinase activity of insulin receptors, 164–5, 225, 381, 382
U
ubiquitination, GLUT4, 177
UK see United Kingdom
cancers, foot see foot
ultrasound
intravascular, atherosclerosis, 1035
skin changes in Rosenbloom syndrome, 959
see also echocardiography
UN and human rights and right to health, 1105, 1106, 1107
uncoupling proteins and obesity treatment, 484–5
United Kingdom, economic burden in, 1115
United Kingdom Prospective Diabetes Study Group (UKPDS), 413, 566, 1094
biguanides, 641, 646
cardiovascular/coronary heart disease, 1012, 1065–6, 1066, 1070, 1075
HbA1c and, 859–60
epiphenetics and, 409
nephropathy, 565, 915, 917, 919, 920
retinopathy? 570, 891, 898
sulfonlureas, 624, 625, 628, 630, 694, 739
United Kingdom Prospective Diabetes Study Group (UKPDS) risk engine, 1073
United Nations (UN) and human rights and right to health, 1105, 1106, 1107
United States (US) and North America
costs paid by individuals and family, 1116–17
Diabetes Prevention Program (DPP), 43, 267, 469, 496, 500, 555, 559, 648, 663, 838, 940
dietary management in, 577–85
T2DM epidemiology, 31
children and adolescents, 40
T2DM prevention, 559
Universal Declaration of Human Rights, 1106
unnanelylated nerve fibers, 930
unsaturated fatty acids (dietary), 579–80
trans-?, 579, 580
see also monounsaturated fatty acids; omega-3 polyunsaturated fatty acids; omega-6 polyunsaturated fatty acids
uremia and HbA1c tests, 862–3
uremic acidosis, laboratory values, 806
urethra, alprostadil pellet in, 984
urinary incontinence, elderly, 841
urogenital autonomic neuropathy, 943, 946–8
US see United States
Uzbekistan, T2DM age and sex-specific prevalence, 32
V
V.A.C. system, 970
vaccines in T1DM
in causation, 22
in treatment, 543
vacuum devices in erectile dysfunction, 984
vaginal and satiety, 508
vaginal vs. cesarean delivery, 832–3
Valeritas V.Go® Insulin Delivery System, 754
Valsalva maneuver in assessment of cardiovascular autonomic neuropathy, 942–3
valsartan, prevention of T2DM, 556
VAMPs (vesicle-associated membrane proteins), 100, 177
vanadium supplements, 720
vapor-permeable dressings, 970
vardenafil, erectile dysfunction, 982, 984
variable number of tandem repeats (VNTR), insulin gene, 85, 445, 460
vascular cell adhesion molecule-1 (VCAM-1), 885
vascular endothelial growth factor (VEGF) and retinopathy
VEGF in pathogenesis, 893
VEGF inhibitors in management, 570–1, 900–1
vascular endothelial growth factor-2 receptor and cardiovascular events, 1047
vascular disease
diabetes and risk of, 1091
subtypes see cardiovascular disease;
cerebrovascular disease; macrovascular disease; microvascular complications; peripheral vascular disease
endothelium see endothelium
foot, assessment, 968
islet, 71–2
penile, and erectile dysfunction, 977
retinal see retinal vessels
smooth muscle cells (VSMCs) and atherogenesis, 665, 1036, 1038
vasoactive intestinal polypeptide (VIP), 127
erective dysfunction and, 976
vasoactive intestinal polypeptide–pituitary adenylate cyclase-activating polypeptide receptors (VPAC), 127, 128, 132
vasodilation, insulin-induced, 251–2
vaspin, 719
VCAM-1 (vascular cell adhesion molecule-1), 885
vegetables
intake, 44
optimal amounts, 584
VEGF see vascular endothelial growth factor; vascular endothelial growth factor-2
ventral pancreatic bud induction, 58
Vero™ system, 753
vertical banded gastroplasty (VBG), 505–6, 508, 509
very-low-carbohydrate diets, 482, 494
very-low-density lipoprotein (VLDL), 262, 263, 264, 289
very-energy diets, 491
vesicle-associated membrane proteins
(VAMP-2/VAMP3), 100, 177
Veterans Administration Genetic Epidemiology Study (VAGES), 373, 687
Veterans Affairs Diabetes Trial (VADT), 309, 565, 566, 878, 898, 1034, 1065, 1066, 1094, 1095
Veterans Affairs Feasibility Trial, 565, 566
Veterans Affairs High-Density Lipoprotein Intervention Trial (VA-HIT), 266, 270
Veterans Affairs Nephropathy-D study, 1085
V-Go® Insulin Delivery System (Valeritas), 754
vildagliptin, 715, 733, 734, 740
pharmacologic characterization, 734
plus DPP-4 inhibitor, 692
plus metformin, 692
structure, 741
viral infections (as cause/in pathogenesis)
diabetes, 6, 9
T2DM, 22, 446–7, 461
T2DM animal studies, 520
obesity, animal studies, 520
viscera (splanchnic tissues)
adiposity, 1032, 1033, 1042
insulin effects on protein metabolism, 255–6
visfatin, 719
vision threatening retinopathy, prevalence, 890, 891
vitamin(s), 720
diabetic recommendations, 583
fat-soluble, colesvelam and deficiency of, 760–1
therapeutic potential, 720
vitamin A receptor activator in nephropathy, 923
vitamin B12 deficiency, metformin-associated, 651
vitamin C and insulin resistance, 345
vitamin D
insulin resistance and, 345
supplementation, 23, 545, 720
vitamin E in nonalcoholic fatty liver disease, 298
vitrectomy, 900, 902
VLBD (very low-density lipoprotein), 262, 263, 264, 289
voglibose, 673, 674, 676
combined with diet therapy, 675
combined with other drugs, 676
prevention of T2DM, 677
von Willebrand factor (vWF), 1055
insulin resistance and, 345
W
waist circumference and abdominal obesity, 348, 349
"Walking Away from Type 2 Diabetes", 559
WASP (Wiskott–Aldrich syndrome proteins), 178
water
doubly-labelled, energy expenditure measurement using, 480
metabolism, in ketoacidosis and hyperosmolar state, 803
retention, sulfonlurea-induced, 626
see also fluid therapy; hydration
weak-acid hydrolysis, total HbA1c measurement, 861
weight (body)
at birth, excessive, 828
fetal, assessment, 831–2
gain
excessive see obesity
with insulin therapy, 794–5
with sulfonlureas, 630–1
with thiazolidinediones, 664, 669
GIP and, 153
loss/reduction (interventional), 267, 489–95
current recommendation in diabetes, 578, 579
maintaining, 495, 496
in nonalcoholic fatty liver disease, 293–4, 296, 297
pharmacological methods, 296, 496–502, 648–9, 678

Index

weight (body) (continued)
prognostic markers of success and reasons for failure, 490–1
surgical methods see bariatric surgery in T2DM prevention, 495, 552, 557
in T2DM treatment, 496
see also diet and nutrition
neutrality of DPP-4 inhibitors, 736–7
well-being, fetal, assessment, 831
Western Samoa, 34, 38
wheat, 22
WHO see World Health Organization
Wisconsin Epidemiologic Study of Diabetic Retinopathy (WESDR), 891, 892, 897
Wisconsin Sleep Cohort Study, 318, 329, 330
Wiskott–Aldrich syndrome proteins (WASPs), 178
Wolcott–Rallison syndrome and EIF2AK3 genes, 427, 434
Wolfram syndrome (DIDMOAD), 7, 427, 435, 437
women
Convention on the Elimination of All Forms of Discrimination against Women, 1107, 1108
coronary heart disease, 1011
acute myocardial infarction fatalities, 1013
postmenopausal, exercise and diabetes 597 females
see women
pregnant see pregnancy
Women's Health Initiative (WHI), 306, 309, 310
World Health Organization (WHO)
diabetes classification, 3, 3–4, 4
diagnostic criteria, 29
epidemic of, 1105
obesity and, 470
HbA1c, 551
hypertension definitions (with ISH), 1007–8
impaired glucose tolerance and fasting glucose, 550, 551
metabolic syndrome diagnostic criteria, 347, 348, 1034
utility of concept, 350–1
Multinational Study of vascular disease in diabetes, 1015, 1097
WOSCOPS, 268
wound care (foot ulcers), 969–70
X
X-ray plain films
Charcot neuroarthropathy, 972
periodontal disease, 989
XENDOS trial, 555, 557
Y
yellow mutation and obesity and cancer, 311
youths/young adults see adolescents and youths/young adults
Z
zebrafish model of endoderm induction and morphogenesis, 57
Zucker fatty rats see fa/fa rats
Figure 3.5 Prevalence of diabetes in men and women in three Arabian countries.

Figure 3.11 Type 2 diabetes, insulin resistance, and puberty in children.
Figure 11.1 The incretin effect on insulin (a) and glucagon (b) secretion. Following oral administration of nutrients, islet hormone secretion is modulated by (A) the incretin hormones GIP and GLP-1, (B) the sympathetic and parasympathetic nervous system, and (C) the circulating substrates glucose, amino acids, and free fatty acids. Stimulatory actions have been denoted with blue color, inhibitory actions are indicated by red color. Source: Adapted from Creutzfeldt 1979 [98]. Reproduced with permission of Springer.
Figure 12.1 Structural and functional domains of the insulin receptor. Schematic diagram of the insulin receptor showing structural landmarks (left side) and functional domains (right side). (Numbering of amino acids based on translation of the alternatively spliced exon 11.)
Figure 12.2 Metabolic and mitogenic signaling pathways in insulin action. Abbreviations: DAG, diacylglycerol; IP3, inositol-3-phosphates; PI(3,4)P2, phosphatidylinositide 3,4 diphosphate; PI(3,4,5)P3, phosphatidylinositide 3,4,5 triphosphate; PI-3 kinase, phosphatidylinositide-3 kinase; PDK1, phosphoinositide-dependent kinase 1; GSK3, glycogen synthase kinase 3; GS, glycogen synthase; GLUT-4, glucose transport protein 4; IRS, family of insulin receptor substrate proteins; SHIP2, SH2-containing inositol phosphatase 2; Grb2, growth factor receptor bound 2; SOS, mammalian homologue of Drosophila son-of-sevenless protein; Akt/PKB, protein kinase B; Shc, adapter protein with homology with Src and collagen; Ras, rat sarcoma protein; Gap, GTPase activating protein; MAP kinase kinase, mitogen activated protein kinase kinase; MAP kinase, mitogen activating protein kinase; PP1G, glycogen-associated protein phosphatase 1.
Figure 12.3 Three major steps providing for functional divergence of insulin signal transduction. The insulin receptor substrate docking proteins, PI-3 kinase, and Akt/PKB are sequentially activated by insulin, but then each factor engages multiple downstream signal molecules and pathways resulting in divergence, capacity for interaction, and complementarity of insulin signaling pathways. Abbreviations: APS, adaptor protein containing PH and SH2 domains; IRS, insulin receptor substrate protein; Gab-1, GRB-2-associated binding protein 1; Shc, adapter protein with homology with Src and collagen; PI(3,4,5)P3, phosphatidylinositide 3,4,5 triphosphate; mTOR, mammalian target of rapamycin; p70S6K, p70 S6 kinase; Akt/PKB, protein kinase B; aPKC, atypical protein kinase C isoforms; GSK3, glycogen synthase kinase-3; BAD, Bcl2 antagonist of cell death; NO, nitric oxide.

Figure 12.4 Two complementary pathways for stimulation of GLUT4 translocation. IRS-1, insulin receptor substrate-1; SHIP2, SH2-containing inositol phosphatase 2; PI-3 kinase, phosphatidylinositide-3 kinase; PKC, protein kinase C; Akt/PKB, protein kinase B; PDK1, phosphoinositide-dependent kinase 1; APS, adaptor protein containing PH and SH2 domains; CAP, c-Cbl associating protein; Cbl, Cbl proto-oncogene; CrkII, Crk oncogene; C3G, guanine nucleotide-releasing protein C3G; TC10, ras-like protein TC10; CIP4/2, Cdc42-interacting protein; GTP, guanosine triphosphate; GDP, guanosine diphosphate.
Figure 12.5 Cellular trafficking and itinerary of GLUT4 glucose transporter proteins. The model is based on a consensus of current data, and includes an endosomal recycling pathway, a separate intracellular inducible GLUT4 storage compartment which supplies the majority of GLUT4 recruited by insulin to the cell surface, and an alternative compartment where accumulated GLUT4 does not respond to insulin in insulin-resistant states (see text).

Figure 12.6 Proteins regulating the docking of exocytotic GLUT4 vesicles at the plasma membrane. Abbreviations: IRAP, insulin regulated aminopeptidase; ACS-1, acyl-CoA synthetase-1; VAMP-2, vesicle-associated membrane protein 2; NSF, N-ethylmaleimide sensitive factor; SNAP-23, soluble NSF attachment 23 kDa protein; Munc-18c, mouse Unc homologue 18c; Synip, syntaxin 4 interacting protein; Rab, monomeric GTPase in the Rab subclass; t-SNARE and v-SNARE refer to target membrane and vesicle-associated soluble N-ethylmaleimide-sensitive attachment protein receptors, respectively.
Figure 12.7  Mechanisms for modulation of insulin signal transduction. The figure illustrates pathways resulting in the desensitization of insulin signal transduction, including serine/threonine phosphorylation of the insulin receptor or IRS by PKC, IKKβ, or JNK; inhibition of the insulin receptor tyrosine kinase by PC-1; tyrosine phosphatase action of PTPase-1B and LAR; catabolism of 3′ phosphoinositide moieties by SHIP-2 or PTEN; inhibition of Akt/PKB activation by ceramide; and interference with substrate binding (STAT5B) to the insulin receptor β subunit by SOCS-3. Abbreviations: PC-1, plasma cell membrane glycoprotein PC-1; SOCS-3, suppressor of cytokine signaling 3; STAT-5B, signal transducer and activator of transcription 5B; PTPase-1B, protein tyrosine phosphatase 1B; LAR, leukocyte antigen-related tyrosine phosphatase; cPKC, classic protein kinase C isoforms; nPKC, novel protein kinase C isoforms; PKC-ζ, protein kinase Cζ; JNK, c-Jun N-terminal kinase; IKK-β, kinase of inhibitor of kappa light chain gene enhancer in B cells β; TNF-α, tumor necrosis factor α; IRS, insulin receptor substrate protein; ser/thr, a serine or threonine residue; Y, tyrosine residue; PI-3 kinase, phosphatidylinositide-3 kinase; SHIP-2, SH2-containing inositol phosphatase 2; PTEN, phosphatase and tensin homologue; Akt/PKB, protein kinase B.
Figure 18.1 Metabolic subdomains of the mammalian metabolome. Analytical protocols using chromatography-coupled mass spectrometry (LC-MS or GC-MS) or proton nuclear magnetic resonance (NMR) are used to simultaneously determine concentrations of metabolites from many metabolic subdomains in urine, blood, or other biofluids. Metabolomics platforms can be either targeted (specific to a metabolite class) or untargeted (more global evaluation of multiple classes in parallel). Source: Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M: Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Research* 2014;42:D199 – D205; Kanehisa M, Goto S: KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* 2000;28:27 – 30.

Figure 22.1 An example of apnea (red), hypopneas (pink), and desaturations (green) in a patient with OSA and T2DM. Note the presence of thoracic and abdominal movements indicating the presence of obstructive rather than central sleep apnea.
Figure 24.7 Preparation of isolated human islets: (a) the pancreas is distended by the injection of a solution containing the digesting enzymes; (b) this leads to the separation of islets (arrows; the red color is due to dithizone staining) from acinar fragments; (c) islet enrichment is achieved by gradient purification, which can be done manually (as in the panel) or automatically; (d) the final islet preparation (dithizone stained) can be used for in vitro studies. Source: Images from PM's laboratory.
Figure 24.9 DPP-4 is present in human islet cells, co-localizing with glucagon (a), but not with insulin (b) or somatostatin (c). Source: Adapted from Omar et al. 2014 [127]. Reproduced with permission of Springer Science and Business Media.

Figure 24.10 The incretin effect, measured as the difference (dark shaded areas) between the insulin secretory response to oral and intravenous (i.v.) glucose under isoglycemic conditions (top panels). Note the absence of a significant incretin effect in patients with type 2 diabetes (T2D). Source: Redrawn from data in Muscelli et al. 2008 [146].
Figure 24.11 Time-course of glucose-induced and incretin-induced potentiation of insulin secretion in nondiabetic subjects (NGT) and in patients with type 2 diabetes. Source: Redrawn from Tura et al. 2014 [135]. Reproduced with permission of Springer Science and Business Media.

Figure 24.13 Schematic representation of the dedifferentiation concept, according to Wajchenberg 2007 [184]. Under normal conditions, β cells have cytoplasmic FOXO1 (yellow); during metabolic stress, insulin production is maintained (green), with FOXO1 translocating to the nucleus (red) to promote β-cell health. With persisting, FOXO1 expression declines (blue nucleus), other transcription factors are reactivated and insulin production decreases (gray). At this stage, former β cells may revert to an uncommitted endocrine progenitor stage (gray) or undergo conversion to cells producing other hormones (orange).
Figure 31.2 Alterations to the composition and metabolic capacity of gut microbiota in obesity promote adiposity and influence metabolic processes in peripheral organs, such as the control of satiety in the brain; the release of hormones from the gut (shown as PYY and GLP-1); and the synthesis, storage or metabolism of lipids in the adipose tissue, liver, and muscle. Microbial molecules also increase intestinal permeability, leading to systemic inflammation and insulin resistance. Source: Tremaroli V, Bäckhed F: *Nature* 2012;489(7415):242–249. Reproduced with permission of Nature Publishing Group.
Figure 34.1 (a) Roux-en-Y gastric bypass: the length of the alimentary and biliopancreatic limbs can vary, but the basic design of this procedure results in nutrients moving directly from gastric pouch to post-duodenum small intestine. The relative importance of the effects of duodenal exclusion and increased nutrient exposure to the ileum remains a matter of debate and a focus of much research. (b) Roux-en-Y gastric bypass: the volume of the gastric pouch formed in this procedure can vary.
Figure 34.2. Adjustable gastric band: the pressure applied through the band can be adjusted by injection or withdrawal of solution through the access port.
Figure 34.3 Sleeve gastrectomy: this procedure does not alter the direction of nutrient flow through the gut.
Figure 34.4 Biliopancreatic diversion: the combination of sleeve gastrectomy and intestinal rearrangement results in significant changes to the metabolism, but can result in malabsorption.
Figure 52.6  Comparison of kidney biopsies of native pancreas in type 1 diabetic recipients of pancreas transplantation before, 5 years, and 10 years post-transplant. The increased basement membrane thickness and increased mesangium characteristic of type 1 diabetes that was present pre-transplant was still present at 5 years post-transplant but greatly decreased at 10 years post-transplant. Source: Adapted from Fioretto 1998 [46]. Reproduced with permission of Massachusetts Medical Society.
**Figure 53.3** Restoration of symptom responses (main panel) to induced controlled hypoglycemia (top panel) before (blue) and after (orange) a period of strict avoidance of hypoglycemia in daily life. Source: Cranston 1994 [37]. Reproduced with permission of Elsevier.

**Figure 53.4** Statistical parametric maps superimposed on MR brain images, showing evolution of changes in regional brain blood flow during experimentally induced hypoglycemia and recovery using water positron emission tomography in healthy volunteers. Blood flow increased (orange) in brain regions involved in stress response, including hypothalamus and pituitary, in symptom perception and awareness including anterior cingulate cortex and in prefrontal regions involved in reward pathways; decreasing (blue) in brain regions involved in balance and memory formation. In the posterior thalamus, flow increased in hypoglycemia and reduced during recovery, the latter perhaps associated with the drowsiness that follows hypoglycemia. Source: Teh et al. [44]. Reproduced with permission of Elsevier.
Figure 63.1 Light microscopic images of sural nerve fascicles from a control subject (a), diabetic patient with minimal (b), moderate (c), and severe (d) neuropathy showing a progressive loss of myelinated fibers (mf) (×300).
Figure 63.7  Light microscopic image of a sural nerve from a diabetic patient with DLSPN showing epineurial vessel pathology (×500).
Figure 63.8  Light microscopic image of a skin biopsy stained with PGP 9.5 showing normal IENF and dermal fibers (a) and a marked depletion of IENF with only dermal NF in a diabetic patient with neuropathy (b).
Figure 66.1 Infected diabetic foot lesion. A 56-year-old patient presented with a painless, necrotic ulcer on the 5th toe of the left foot. On examination there was surrounding erythema and purulent discharge clinically suggesting infection.
Figure 66.2 Pressure-induced plantar ulceration. This 45-year-old diabetic patient with insensitive feet presented with shoe-induced ulcers on the toes of both feet. The hallux plantar surface showed build-up of callus and there were necrotic lesions on the apices of the 2nd and 3rd toe of the right foot.

Figure 66.3 Thermal foot wound. This 48-year-old female diabetic patient with peripheral neuropathy presented with extensive burns to the plantar surface of the metatarsal head region and hallux caused by walking on hot sand whilst on holiday.
Figure 69.1 Cumulative survival curves for incidence of coronary heart disease (a) and for cardiovascular mortality (b) according to the glucose categories defined by both fasting and 2-h plasma glucose criteria. The cumulative survival are estimated from Cox proportional hazards models and adjusted for age, cohorts, sex, body mass index, blood pressure, serum cholesterol, and smoking. Normal, fasting plasma glucose <6.1 mmol L⁻¹ and 2-h plasma glucose <7.8 mmol L⁻¹; IFG, impaired fasting glycemia only; IGT, impaired glucose tolerance only; IGR, impaired glucose regulation (IFG and IGT); DMF, undiagnosed diabetes by fasting glucose criteria alone (fasting plasma glucose >7.0 mmol L⁻¹ and 2-h plasma glucose <11.1 mmol L⁻¹); DMP, undiagnosed diabetes by 2-h postload glucose criteria alone (2-h plasma glucose ≥11.1 mmol L⁻¹ and fasting plasma glucose <7.0 mmol L⁻¹); DMC, undiagnosed diabetes by both fasting and 2-h glucose criteria combined (fasting plasma glucose ≥7.0 mmol L⁻¹ and 2-h plasma glucose ≥11.1 mmol L⁻¹); known DM, previously diagnosed diabetes.
WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley’s ebook EULA.