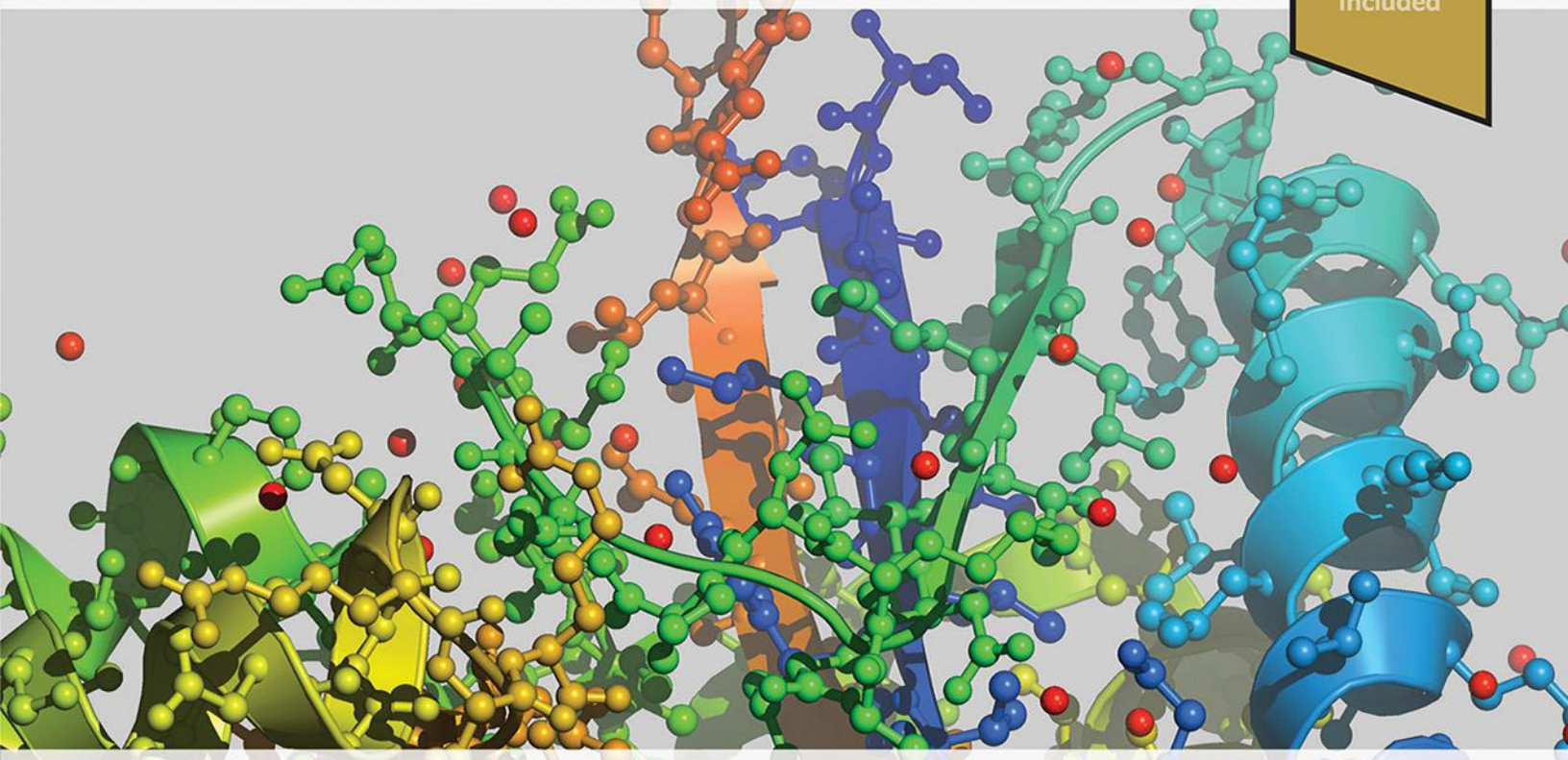
JOHN W. BAYNES
MAREK H. DOMINICZAK

Enhanced DIGITAL VERSION Included



SIXTH EDITION

Medical BIOCHEMISTRY

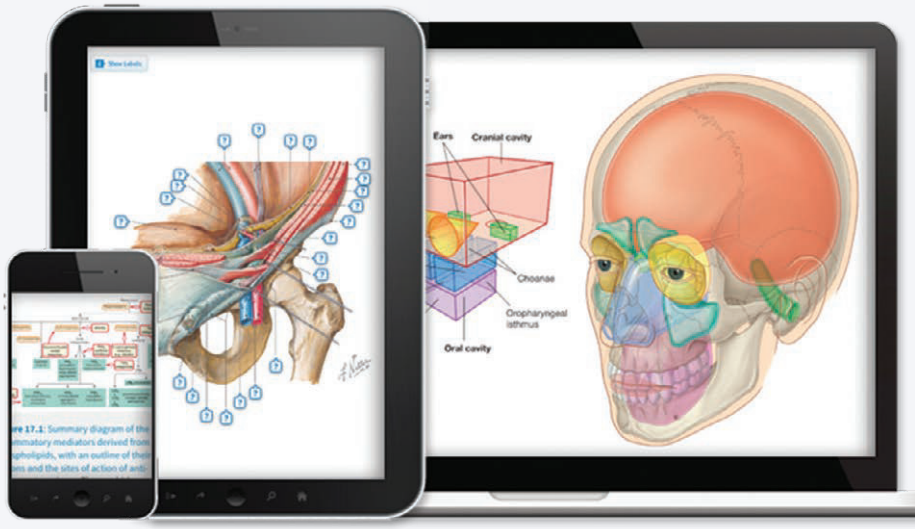
ELSEVIER

eBooks for Medical Education

ELSEVIER

Any screen. Any time. Anywhere.

Activate the eBook version
of this title at no additional charge.



Elsevier eBooks for Medical Education gives you the power to browse and search content, view enhanced images, highlight and take notes—both online and offline.

Unlock your eBook today.

1. Visit **studentconsult.inkling.com/redeem**
2. Scratch box below to reveal your code
3. Type code into "Enter Code" box
4. Click "Redeem"
5. Log in or Sign up
6. Go to "My Library"

It's that easy!

Place Peel Off Sticker Here

For technical assistance:
email studentconsult.help@elsevier.com
call 1-800-401-9962 (inside the US)
call +1-314-447-8300 (outside the US)

2020_ME

SIXTH EDITION

Medical BIOCHEMISTRY

Caption: Resident2. Pencil on paper by Marek H. Dominiczak©

The next step in your career in medicine is your residency, where you will be in a position to help patients by solving clinical problems. This will require you to make decisions on diagnosis and treatment. The reason you learn basic science, including biochemistry, is to hone your clinical thinking so that these decisions are better informed and effective.

We have placed this sketch by Marek Dominiczak here to remind ourselves that one should always see learning biochemistry in the context of this future role.

SIXTH EDITION

Medical BIOCHEMISTRY

JOHN W. BAYNES, PhD

Carolina Distinguished Professor Emeritus

Department of Pharmacology, Physiology and Neuroscience
University of South Carolina School of Medicine
Columbia, SC, USA

MAREK H. DOMINICZAK, MD, Dr Hab Med, FRCPath, FRCP (Glas)

Hon Professor of Clinical Biochemistry and Medical Humanities

College of Medical, Veterinary and Life Sciences
University of Glasgow
Glasgow, UK

Consultant Biochemist Emeritus

Clinical Biochemistry Service
National Health Service (NHS) Greater Glasgow and Clyde
Glasgow, UK

ELSEVIER

ELSEVIER

First edition 1999
Second edition 2005
Third edition 2009
Fourth edition 2014
Fifth edition 2019
Sixth Edition 2022

Notices

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds or experiments described herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made. To the fullest extent of the law, no responsibility is assumed by Elsevier, authors, editors or contributors for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

ISBN: 978-0-323-83450-6

Content Strategist: Jeremy Bowes
Content Development Specialist: Rishi Arora
Manager: Sreejith Viswanathan
Design: Ryan Cook
Illustration Manager: Karen Giacomucci
Marketing Manager: Belinda Tudin

Printed in China
Last digit is the print number: 9 8 7 6 5 4 3 2 1

Contents

List of Contributors

Catherine Bagot, BSc, MBBS, MD, MRCP, FRCPath
Consultant Hematologist
Department of Haematology
Glasgow Royal Infirmary
Honorary Senior Lecturer
Institute of Cardiovascular and Medical Sciences
University of Glasgow, Glasgow
United Kingdom

John W. Baynes, PhD
Carolina Distinguished Professor Emeritus
Pharmacology, Physiology and Neuroscience
University of South Carolina School of Medicine
Columbia
South Carolina
United States

Sophie Bradley, BSc, PhD
Hon Senior Research Fellow
Institute of Molecular Cell and Systems Biology
University of Glasgow, Glasgow
United Kingdom

Iain Broom MBChB, MIBiol, FRCPath
Professor Emeritus of Metabolic Medicine
University of Aberdeen
Aberdeen
United Kingdom

Wayne E. Carver, PhD
Professor and Chair
Cell Biology and Anatomy
University of South Carolina School of Medicine
Columbia
South Carolina
United States

Norma Frizzell, BSc, PhD
Associate Professor
Pharmacology, Physiology and Neuroscience
University of South Carolina School of Medicine
Columbia
South Carolina
United States

Marek Dominiczak, MD, Dr Hab Med, PhD, FRCPath, FRCP (Glas)
Honorary Professor
College of Medical Veterinary and Life Sciences
University of Glasgow
Consultant Clinical Biochemist Emeritus
National Health Service (NHS) Greater Glasgow
and Clyde
Glasgow
United Kingdom

William Fraser BSc, MBChB, MD, FRCP, FRCPath
Professor
Consultant Clinical Chair in Medicine
Norwich Medical School
Faculty of Medicine and Health Sciences
Norwich
United Kingdom

Junichi Fujii, PhD
Professor
Biochemistry and Molecular Biology
Graduate School of Medical Science, Yamagata University
Yamagata
Japan

Alastair Gracie, PhD, BSc
Professor, Honorary Senior Research Fellow (Medicine)
School of Medicine, Dentistry and Nursing
University of Glasgow
Glasgow
United Kingdom

Alejandro Gugliucci, MD, PhD
Professor Emeritus of Biochemistry
Touro University California, Vallejo
California
United States

Margaret Harnett, PhD
Professor of Immune Signalling
School of Life Sciences
University of Glasgow
Glasgow
United Kingdom

Simon Heales, PhD, FRCPath
Professor of Clinical Biochemistry
Department of Laboratory Medicine
Great Ormond Street Hospital
London
United Kingdom

Koichi Honke, MD, PhD
Professor
Department of Biochemistry
Kochi University Medical School, Nankoku
Kochi
Japan

Edel M. Hyland, PhD
Senior Lecturer
School of Biological Science
Queens University Belfast
Antrim, Northern Ireland
United Kingdom

Susan Johnston, BSc, MSc, FRCPath
Clinical Biochemist
Clinical Biochemistry Service
National Health Service (NHS) Greater Glasgow and Clyde
Glasgow
United Kingdom

Alan F Jones, MA, MB, BChir, DPhil, FRCP, FRCPath
Consultant Physician
Heart of England NHS Foundation Trust
Bordesley Green East
Birmingham
United Kingdom

Fredrik Karpe, MD, PhD
Professor of Metabolic Medicine
Oxford Centre for Diabetes, Endocrinology and Metabolism
University of Oxford
Oxford
United Kingdom

Walter Kolch, MD
Professor
Systems Biology Ireland
University College Dublin, Dublin
Ireland

Matthew C. Kostek, BS, MS, PhD
Associate Professor
Physical Therapy
Duquesne University, Pittsburgh
Pennsylvania
United States

Associate Professor
McGowan Institute of Regenerative Medicine
University of Pittsburgh School of Medicine, Pittsburgh
Pennsylvania
United States

Jennifer Logue, MbChB, MRCP, MD, FRCPath
Professor and Honorary Consultant in Metabolic Medicine
Lancaster Medical School
University of Lancaster
Lancaster
United Kingdom

Masatomo Maeda, PhD
Professor Emeritus
Graduate School of Pharmaceutical Sciences
Osaka University, Suita
Japan
Professor Emeritus
School of Pharmacy
Iwate Medical University, Morioka
Iwate
Japan

Teresita Menini, MD, MS
Professor Emerita of Biochemistry
Touro University California, Vallejo
California
United States

Alison M. Michie, PhD, BSc
Reader in Molecular Lymphopoiesis
Institute of Cancer Sciences
University of Glasgow
Glasgow
United Kingdom

Ryoji Nagai, PhD
Professor
Laboratory of Food and Regulation Biology
Tokai University
Kumamoto
Japan

Verica Paunovic, PhD
Senior Research Associate
Institute of Microbiology and Immunology
Faculty of Medicine
Belgrade
Serbia

Georgia Perona-Wright, PhD
Senior Lecturer
Institute of Infection, Immunity and Inflammation
University of Glasgow, Glasgow
United Kingdom

Andrew R. Pitt, PhD
Professor of Pharmaceutical Chemistry
and Chemical Biology
University of Manchester
England,
United Kingdom

Simon Pope, PhD
Clinical Biochemist
Neurometabolic Unit
National Hospital
UCLH Foundation Trust
London
United Kingdom

Matthew Priest, MB, ChB, FRCP (Glas)
Consultant Gastroenterologist
National Health Service (NHS) Greater Glasgow and Clyde
Honorary Senior Lecturer
University of Glasgow
Glasgow
United Kingdom

Allen B. Rawitch†, AA, BS, PhD
Professor Emeritus
Department of Biochemistry and Molecular Biology
University of Kansas Medical Center, Kansas City
Kansas
United States
Vice Chancellor Emeritus
University of Kansas Medical Center, Kansas City
Kansas
United States

Ian Salt, PhD
Senior Lecturer
Institute of Cardiovascular and Medical Sciences
University of Glasgow
Glasgow
United Kingdom

Yee Ping Teoh, MBBS, MRCP, FRCPath
Consultant in Chemical Pathology and Metabolic Medicine
Clinical Biochemistry
Wrexham Maelor Hospital
Wrexham
United Kingdom

Robert Thornburg, BS, PhD
Professor Emeritus
Department of Biochemistry, Biophysics
and Molecular Biology
Iowa State University
Ames, Iowa
United States

†Author deceased

This page intentionally left blank

Acknowledgments

First of all, we wish to thank our contributors for sharing their expertise with us and for fitting the writing—again—into their busy research, teaching, and clinical schedules. Most of the preparation of this edition took place during the wordwide COVID-19 pandemic, and this put additional strain on people's professional and personal lives. We are most grateful for their resilience.

In the sixth edition, we welcome new contributors Georgia Perona-Wright and Sophie Bradley.

Our inspiration to change and improve this text comes from the problems, questions, and decisions that arise in our everyday clinical practice, in the outpatient clinics, and on the hospital wards. We are grateful to all our clinical colleagues and doctors in training for their insight, discussions, and sharing of their clinical experience. We are also grateful to students and academics from universities around the world who continue to provide us with comments, suggestions, and criticisms. We acknowledge the contribution of scholars who participated in the writing of previous editions of *Medical Biochemistry:* Gary A. Bannon, Graham Beastall, Robert Best, Franklin F. Bolander, James A. Carson, Alan D. Elbein (deceased), Alex Farrell (deceased), Helen S. Goodridge, George M. Helmkamp Jr., D. Margaret Hunt, Andrew Jamieson, Gur P. Kaushal, W. Stephen Kistler, Utkarsh V. Kulkarni, Jeffrey R. Patton, L. William Stillway, Naoyuki Taniguchi, and A. Michael Wallace (deceased).

We remain very grateful to Jacky Gardiner in Glasgow, now retired, who provided excellent secretarial assistance throughout five previous editions of this book.

The key to the success of the entire project has been, of course, the Elsevier team. Our particular thanks go to Nani Clansey, Louise Cook, Beula Christopher and Jeremy Bowes, who initiated the sixth edition. We thank Jeremy for guiding the project throughout. Our thanks go also to Rishi Arora, Poulouse Joseph, and Sreejith Vishwanathan, who gave the book its final form. Last but not least, we thank designers and illustrators for their effort and patience.

This page intentionally left blank

To inspirational academics

Inquisitive students

And all those who want to be good doctors

This page intentionally left blank

Preface

Medical Biochemistry has now served the global medical student community for 23 years. In this sixth edition, our aim continues to be to provide a biochemical foundation for the study of clinical medicine—with emphasis on down-to-earth practical relevance.

Each edition of a textbook like ours provides a snapshot of a constantly changing field. In biochemistry, the striking sign of progress has been the ever-increasing relevance of basic science to the practice of medicine. What once might have appeared as leading-edge theoretical science is after a few years described as part of routine clinical practice. In particular, there have been more and more drugs targeting specific regulatory pathways, and a growing range of therapies that act at the level of RNA and DNA. We also see how application of many discoveries affects the health of populations. Novel therapies for prevention and treatment of diabetes, atherosclerosis and cardiovascular disease, cancer, and genetic diseases are good examples here.

In addition, the elucidation of mechanisms of disease has led to a greater convergence of disciplines: for example, we now talk about atherosclerosis, diabetes, and nutrition in a much more integrated way than only a few years ago. Throughout the sixth edition of *Medical Biochemistry*, we try to address this evolution. Thus, the core of this book is the account of established current knowledge. Biochemistry of specialized tissues is addressed in later chapters. The emerging findings are addressed throughout in advanced concept boxes, and practical applications are illustrated in clinical test boxes and through clinical cases. We also address the evolving integration of disciplines by multiple cross-references throughout the chapters (see Figure 1.4 in Chapter 1).

We emphasize the contribution of biochemistry to the understanding of global public health problems, such as diabetes mellitus, obesity, malnutrition, and atherosclerotic cardiovascular disease. We also remain convinced that knowledge of water, electrolytes, and acid–base balance is as important for future clinicians as is the knowledge of key metabolic pathways. Thus, *Medical Biochemistry* puts more emphasis on these topics than most other textbooks.

In the previous (fifth) edition, we changed the structure of the book to provide a clearer perspective on the entire field. The details of this are summarized in Chapter 1. In the sixth edition, we have updated literature and particularly web references throughout. To facilitate familiarity with new terminologies and acronyms, an easily accessible list of abbreviations is given at the end of each chapter.

In addition, a question bank for self-assessment and many more resources are available at the Elsevier website, https://studentconsult.com, to which the reader is referred. There is also a companion publication, *Medical Biochemistry Flash Cards*, which provides a means for quick review.

As before, we welcome comments, criticisms, and suggestions from our readers. There is no better way to continue making this a better text.

This page intentionally left blank

Abbreviations

$1,25(OH)_2D_3$	1,25-dihydroxycholecalciferol, calcitriol
$25(OH)D_3$	25-hydroxycholecalciferol, calcidiol
1,3-BPG	1,3-bisphosphoglycerate
17-OHP	17-hydroxyprogesterone
2,3-BPG	2,3-bisphosphoglycerate
4E-BP1	eIF4E-binding protein 1
5-ALA	5-aminolevulinate
5-HIAA	5-hydroxyindoleacetic acid
5-HT	5-hydroxytryptamine, serotonin
8-oxoG	8-oxo-2′-deoxyguanosine
α-MSH	Melanocortin
A1AT	α-1 antitrypsin
AADC	Aromatic amino acid decarboxylase
ABC	ATP binding cassette
ABCA1, ABCG5, G8, A1, G1, and G4	ATP-binding cassette transporters
ABL	Nonreceptor protein tyrosine kinase
ACAT	Acyl-CoA: acyl-cholesterol transferase
ACAT	Cholesterol acyltransferase
ACC1, ACC2	Acetyl-CoA carboxylase
ACD	Autophagic cell death (autophagy)
ACE	Angiotensin-converting enzyme
Acetyl-CoA	Acetyl-coenzyme A
ACh	Acetylcholine
ACP	Acyl carrier protein
ACTH	Adrenocorticotropic hormone
AD	Alzheimer's disease
ADA	American Diabetes Association
ADAR	Adenosine deaminase acting on RNA
ADH	Alcohol dehydrogenase
ADH	Antidiuretic hormone, vasopressin
AE	Anion exchanger
AFP	α-fetoprotein
AG	Anion gap
AGE	Advanced glycation (glycoxidation) end products
AGPAT2	Acylglycerol acyltransferase 2
AHA	American Heart Association
AHF	Antihemophilic factor
AI	Adequate Intake
AIC	Acute intermittent porphyria
AKT	A protein kinase
ALD	Alcoholic liver disease
ALDH	Aldehyde dehydrogenase
ALE	Advanced lipoxidation end-products
ALL	Acute lymphoblastic leukemia
ALP	Alkaline phosphatase
ALPS	Autoimmune lymphoproliferative syndrome
ALT	Alanine aminotransferase
AML	Acute myeloid leukemia
AMPA	α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid
AMPK	AMP-activated protein kinase; AMP-dependent protein kinase
ANP	Atrial natriuretic peptide
AP-1	Activator protein-1
APAF1	Apoptotic protease activating factor 1
APC	Anaphase-promoting complex
APC	Antigen-presenting cell
apoA	Apolipoprotein A
apoB	Apolipoprotein B
apoB100/apoB48	Variants of apolipoprotein B
apoC	Apolipoprotein C
apoE	Apolipoprotein E
APP	Amyloid precursor protein
APRT	Adenosine phosphoribosyl transferase
APTT	Activated partial thromboplastin time
AQP	Aquaporin
ARDS	Acute respiratory distress syndrome
ARE	Antioxidant response element
ASCVD	Atherosclerotic cardiovascular disease
AST	Aspartate aminotransferase
AT1, AT2	Angiotensin receptors type 1 and type 2
ATCase	Aspartate transcarbamoylase
ATF	Activation transcription factor
ATG	Autophagy-related gene
ATM	Ataxia-telangiectasia mutated, checkpoint kinase
ATP III	National Cholesterol Education Treatment Panel III
ATP	Adenosine triphosphate
ATR	Ataxia-telangiectasia Rad3–related checkpoint kinase, CHK1 and CHK2
AUC	Area under the curve
AVP	Arginine vasopressin
AZT	Azidothymidine
BAD	Bcl-2-associated death promoter
Bak	Bcl-2 homologous antagonist/killer
BAX	Bcl-2-associated X protein
BBB	Blood–brain barrier

Bcl-2	B-cell lymphoma protein 2; Bcl-2 family members include prosurvival family members (Bcl-2, Bcl-xL, Bcl-W, Mcl-1); proapoptotic BAX/BAK family, and proapoptotic BH-3-only proteins (BIM, Bid, PUMA, NOXA, BAD, BIK)
BCR	B-cell receptor
BCR	Breakpoint cluster region
BH-3	Interacting-domain death agonist
BH4	Tetrahydrobiopterin
BMI	Body mass index
BMR	Basal metabolic rate
BNP	Brain natriuretic peptide
BrdU	Bromodeoxyuridine
Btk	A protein tyrosine kinase
BUN	Blood urea nitrogen
bw	Body weight
C1q, C1r, C1s, and C2–C9	Complement components
C3G	Guanyl nucleotide exchange factor
CA	Carbonic anhydrase
CAD	Carbamoyl phosphate synthetase-aspartate transcarbamoylase-dihydroorotase
CAH	Congenital adrenal hyperplasia
CAK	CDK-activating complex, composed of CDK7, cyclin H, and MAT1 (ménage a trois)
CaM	Calmodulin
cAMP	Adenosine 3′,5′-cyclic monophosphate
cAMP	Cyclic adenosine monophosphate
CAMS	Cell adhesion molecules
CAP	Cbl-associated protein
CAT	Catalase
CBG	Cortisol-binding globulin (also known as transcortin)
Cbl	Adaptor protein in insulin signaling pathway
CD	Cluster of differentiation system; cell surface molecules
$CD4^+$	T helper cells (T_H)
CD40L	CD40 ligand
$CD8^+$	Cytotoxic T lymphocyte (CTL)
CDG	Congenital disease of glycosylation
CDK	Cyclin-dependent kinase
CDKI	Cyclin-dependent kinase inhibitory protein
cDNA	Complementary DNA
CDP	Cytidine diphosphate
CDP-DAG	CDP-diacylglycerol
CE	Cholesterol ester
CEA	Carcinoembryonic antigen
CETP	Cholesterol ester transfer protein
CF	Cystic fibrosis
CFDA SE	Carboxyfluorescein diacetate succinimidyl ester
cFLIP	Modulator of FADD, Fas-associated death domain
CFTR	Cystic fibrosis transmembrane conductance regulator
CGD	Chronic granulomatous disease
CGH	Comparative genome hybridization
cGMP	Cyclic guanosine 3′5′-monophosphate
cGMP	Cyclic guanosine monophosphate
C_H	Constant heavy fragment; antigen-binding sequence domains
ChAT	Choline acetyltransferase
ChIP	Chromatin immunoprecipitation
ChIP-on-chip	Combination chromatin immunoprecipitation and microarray technology
ChIPseq	Combination chromatin immunoprecipitation and RNAseq technology
CHK1	CHK2, checkpoint kinases
CK	Creatine (phospho)kinase
CK-MB	MB fraction of creatine kinase
C_L	Constant light, fragment; antigen-binding sequence domains
CLL	Chronic lymphocytic leukemia
CLR	C-type lectin receptors
CMA	Chromosomal microarray analysis
CML	Chronic myeloid leukemia
CML	Nε-(carboxymethyl)lysine
CMP-NeuAc	CDP-neuraminic (sialic) acid
CMP-PA	Cytosine monophosphate-phosphatidic acid
CNS	Central nervous system
COAD	Chronic obstructive airway disease
CoA-SH	Acetyl-coenzyme A
COHb	Carboxyhemoglobin
COMT	Catecholamine-O-methyltransferase
CpG	Cystine-guanine dinucleotide
CPS	Carbamoyl phosphate synthetase
CPT-I, CPT-II	Carnitine palmitoyl transferase I and II
CRBP	Cytosolic retinol-binding proteins
CREB	cAMP response element-binding protein
CRH	Corticotropin-releasing hormone; corticoliberin
cRNA	Complementary RNA
CRP	C-reactive protein
CSC	Cancer stem cell
CSF	Cerebrospinal fluid
CT	Computed (computerized) tomography scan
CTD	C-terminal domain
CTL	Cytotoxic T lymphocytes ($CD8^+$ cells)
CTX	Carboxy-terminal telopeptide
Cyt a, b, c	Cytochrome a, cytochrome b, cytochrome c
DAG	Diacylglycerol
DAMPs	Damage-associated molecular patterns
DAPI	4′-6′-diamidino-2-phenylindole
DAT	Dopamine transporter
DC	Dendritic cell
DCCT	Diabetes Control and Complications Trial
DD	Death domain
DDI	Drug–drug interaction

DED	Death effector domain
DEXA	Dual-energy x-ray absorptiometry
DGAT	Diacylglycerol acyltransferase
DHAP	Dihydroxyacetone phosphate
DHEA	Dehydroepiandrosterone
DHEAS	Dehydroepiandrosterone sulfate
DHT	Dihydrotestosterone
DIC	Disseminated intravascular coagulation
DILI	Drug-induced liver injury
DISC	Death-inducing signaling complex
DIT	Diiodotyrosine
DLDH	Dihydrolipoyl dehydrogenase
DLTA	Dihydrolipoyl transacetylase
DMP	Dentin matrix protein
DNA	Deoxyribonucleic acid
DNL	*De novo* lipogenesis
DNP	Dinitrophenol
Dol	Dolichol
DPP-4	Dipeptidyl peptidase-4
DPPC	Dipalmitoylphosphatidylcholine
DRI	Dietary Reference Intakes
dsRNA	Double-stranded RNA
DTI	Direct thrombin inhibitor
DVT	Deep vein thrombosis
E2F	Family of transcription factors
EAR	Estimated Average Requirement
EBNA1	Epstein–Barr virus nuclear antigen 1
ECF	Extracellular fluid
ECM	Extracellular matrix
EDRF	Endothelium-derived relaxing factor (nitric oxide)
EDTA	Ethylenediaminetetraacetic acid
eEF	Eukaryotic elongation factor
EFA	Essential fatty acid
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
eGFR	Estimated glomerular filtration rate
eIF	Eukaryotic initiation factor
EMSA	Electrophoretic mobility shift assay
ENaC	Amiloride-sensitive calcium channel
ENaC	Epithelial sodium channel
eNOS	Endothelial nitric oxide synthase
EPA	Eicosapentaenoic acid
Epacs	Exchange proteins directly activated by cAMP
ER	Endoplasmic reticulum
ERAD	ER-associated degradation pathway
eRF	Eukaryotic releasing factor complex
ERK 1 and 2	Extracellular-signal regulated kinases; two isoforms of MEK kinase that activate MAPK
ESR	Erythrocyte sedimentation rate
ETC	Electron transport chain
Fab	Fragment antigen-binding
FAD/$FADH_2$	Flavin adenine dinucleotide (oxidized/reduced)
FADD	Fas-associated death domain (Fas, death receptor, TNF family member)
FasL	Fas ligand
Fc	"Fragment constant" of immunoglobulin molecule
FcγR	Fc-γ receptor (receptor for immunoglobulin G)
FDB	Familial defective apolipoprotein B
FDPs	Fibrin degradation products
FGF	Fibroblast growth factor
FGFR3	Fibroblast growth factor receptor 3
FH	Familial hypercholesterolemia
FIRKO	Adipose tissue (fat) insulin receptor knockout
FISH	Fluorescence in situ hybridization
FMN	Flavin mononucleotide
FOXA2	Transcription factor, also known as HNF-3B
FOXO	Forkhead box O proteins; transcription factors belonging to the forkhead family (contain proteins designated FOXA to FOXR)
FOXP3	Transcription factor
FP	Flavoprotein
FRTA	Free radical theory of aging
Fru-1,6-BP	Fructose-1,6-bisphosphate
Fru-1,6-BPase	Fructose-1,6-bisphosphatase
Fru-1-P	Fructose-1-phosphate
Fru-2,6-BP	Fructose-2,6-biphosphate
Fru-2,6-BPase	Fructose-2,6-bisphosphatase
Fru-6-P	Fructose-6-phosphate
FSF	Fibrin-stabilizing factor
FSH	Follicle-stimulating hormone
fT3 and fT4	Free T3 and free T4
FVII	Factor VII
FXR	Farnesyl X receptor
FYN	Nonreceptor protein tyrosine kinase
G0	Resting, or quiescent, phase of cell cycle
G1	Interval between M and S phases
G2	Interval between S and M phases
G6PDH	Glucose-6-phosphate dehydrogenase
GABA	γ-Aminobutyric acid
GAD	Glutamic acid decarboxylase
GAG	Glycosaminoglycan
Gal	Galactose
Gal-1-P	Galactose-1-phosphate
GALD-3-P	Glyceraldehyde-3-phosphate
GalNAc	N-acetylgalactosamine
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GAPs	GTPase-activating protein
GAS	Gamma interferon activation site
GC-MS	Gas chromatography–mass spectrometry
GCS	Glycine cleavage systems
GDM	Gestational diabetes mellitus
GDP-Fuc	Guanosine diphosphate fucose
GDP-Man	Guanosine diphosphate mannose
GFAP	Glial fibrillary acidic protein
GFR	Glomerular filtration rate

GGT	γ-Glutamyl transpeptidase
GH	Growth hormone
GHRH	Growth hormone–releasing hormone
GI	Gastrointestinal (tract)
GI	Glycemic index
GIP	Gastric inhibitory peptide
GK	Glucokinase
Glc	Glucose
Glc-6-P	Glucose-6-phosphate
Glc-6-Pase	Glucose-6-phosphatase
GlcNAc	N-acetylglucosamine
$GlcNH_2$	Glucosamine
GlcUA	Glucuronic acid
GLP-1	Glucagon-like peptide-1
GLUT	Glucose transporter
Glycerol-3-P	Glycerol-3-phosphate
GnRH	Gonadotropin-releasing hormone
GPCR	G-protein coupled receptor
GPI	Glycosylphosphatidylinositol anchor
GPIb-IX, GPIIb-IIIa	Platelet membrane glycoprotein receptors
GPx	Glutathione peroxidase
GRB2	Growth factor receptor-bound protein 2, adapter molecule
GSH	Glutathione (reduced)
GSSG	Glutathione (oxidized)
GTPase	Guanosine triphosphatase
GWAS	Genome-wide association study
Hb	Hemoglobin
HbA	Adult (normal) hemoglobin
HbA_{1c}	Hemoglobin A_{1c}, glycated hemoglobin
HbF	Fetal hemoglobin
HbS	Sickle cell hemoglobin
hCG	Human chorionic gonadotrophin
HCL	Hairy cell leukemia
hCS-A, hCS-B, hCS-L, and *hGH-V*	Human somatomammotropin (GH) genes
HDL	High-density lipoprotein(s)
HFE	Hereditary hemochromatosis protein
HGF	Hepatocyte growth factor
hGH	Human growth hormone
HGP	Human genome project
HGPRT	Hypoxanthine-guanine phosphoribosyl transferase
HIT	Heparin-induced thrombocytopenia
HIV	Human immunodeficiency virus
HLA	Human leukocyte antigen
HLA-DR, HLA-DQ, HLA-DM, and HLA-DP	MHC class II genes
HMDB	Human Metabolome Database
HMG-CoA	3-hydroxy-3-methylglutaryl-CoA
HMGR	HMG-CoA reductase
HMWK	High-molecular-weight kininogen
HNE	Hydroxynonenal
HNF1A, HNF1B	Transcription factors
hnRNA	Heterogeneous nuclear RNA
HPLC	High-performance liquid chromatography
HRE	Hormone response element
HRG	Histidine-rich glycoprotein
HSP	Heat shock protein
HSV	Herpes simplex virus
HTGL	Hepatic triglyceride lipase
HVA	Homovanillic acid
IAP	Inhibitor of apoptosis gene family
ICAM-1	Intercellular adhesion molecule 1 (CD54)
ICF	Intracellular fluid
IDDM	Insulin-dependent diabetes mellitus
IDL	Intermediate-density lipoprotein(s)
IdUA	Iduronic acid
IEF	Isoelectric focusing
IF	Intrinsic factor
IFCC	International Federation of Clinical Chemistry and Laboratory Medicine
IFG	Impaired fasting glucose
IFN	Interferon (IFN-α, IFN-β, and IFNγ)
Ig	Immunoglobulin (IgG, IgA, IgM, IgD, and IgE)
IGF	Insulin-like growth factor
IGFBP	IGF-binding proteins
IGT	Impaired glucose tolerance
Ihh	Indian hedgehog, a signaling protein
IKK	NFκB kinase
IL	Interleukin (IL-1, IL-6, etc.)
IMAC	Immobilized metal affinity chromatography
IMM	Inner mitochondrial membrane
IMP	Inosine monophosphate
IMS	Intermembrane space
INR	International Normalized Ratio
IP_3	Inositol 1,4,5-trisphosphate
IPP	Isopentenyl diphosphate
IR	Insulin receptor
IRE	Iron response element
IRES	Internal ribosomal entry site
IRI	Ischemic reperfusion injury
IRS	Insulin receptor substrate
ITAM/ITIM	Immunoreceptor tyrosine activation/inhibition motif
IU	International unit
JAK	Janus kinase
JAK/STAT	Janus kinase/signal transducer and activator of transcription
JNK	C-Jun terminal kinase
kb	Kilobase
KCC1	K^+ and Cl^- cotransporter
KCCT	Kaolin–cephalin clotting time, APTT
KIP2	57-kDa inhibitor of cyclin–CDK complexes
KIT	Tyrosine kinase 3 genes
KLF	Kruppel-like factor

K_m	Michaelis constant
LACI	Lipoprotein-associated coagulation inhibitor
LBBB	Left bundle branch block
LC3	Microtubule-associated protein light chain 3
LCAT	Lecithin:cholesterol acyltransferase
LC-MS	Liquid chromatography–mass spectrometry
LDH	Lactate dehydrogenase
LDL	Low-density lipoprotein(s)
LFA-1	Lymphocyte function–associated antigen 1
LH	Luteinizing hormone
LMWH	Low-molecular-weight heparin
lncRNA	Long noncoding RNA
LPL	Lipoprotein lipase
LPLAT	Lysophospholipid acyltransferase
LPS	Lipopolysaccharide
LRP5	LDL-receptor-related protein 5
LSC	Laser-scanning cytometry
LT	Leukotriene
LTA	Light transmission aggregometry
LXR	Liver X receptors
M	Mitosis
MAC-1	Macrophage adhesion molecule 1
MAG	Monoacylglycerol
MALT	Mucosa-associated lymphoid tissues
Man-6-P	Mannose-6-phosphate
MAO	Monoamine oxidase
MAOI	Monoamine oxidase inhibitor
MAPK	Mitogen-activated protein kinase
MAS	Angiotensin 1–7 receptor
Mb	Myoglobin
MBL	Mannose-binding lectin
MCH	Melanin-concentrating hormone
MCL	Mantle cell lymphoma
MCP	Multicatalytic protease
MCP-1	Monocyte chemoattractant protein 1
M-CSF	Monocyte-colony stimulating factor
MCV	Mean corpuscular volume
MDA	Malondialdehyde
MDRD	Modification of diet in renal disease study
MEK	Mitogen-activated protein kinase. A protein kinase that activates MAPK
MET	Metabolic equivalent of task
metHb	Methemoglobin (Fe^{3+})
MetSO	Methionine sulfoxide
MGO	Methylglyoxal
MGP	Matrix gla protein
MHC	Major histocompatibility complex
miRNA	MicroRNA
MIT	Monoiodotyrosine
MMP	Matrix metalloproteinase
MMP	Mitochondrial membrane potential
MODY	Maturity-onset diabetes of the young
MPO	Myeloperoxidase
MRI	Magnetic resonance imaging
MRM	Multiple reaction monitoring
mRNA	Messenger RNA
MRP	Multidrug resistance–associated protein
MS	Mass spectrometry
MSH	Melanocyte-stimulating hormone
MSLP	Maximum lifespan potential
mtDNA	Mitochondrial DNA
MTHFR	5,10-methylenetetrahydrofolate reductase
mTOR	Mechanistic target of rapamycin; a serine/threonine protein kinase
mTORC-1 and mTORC-2	mTor complexes
mTORC	Mammalian target of rapamycin complex
MTP	Microsomal transfer protein
MudPIT	Multidimensional protein identification technology
MWCO	Molecular weight cut-off
Myc	Transcription factor
N5MeTHF	5Methyl tetrahydrofolate
N^5-N^{10}-THF	N^5-N^{10}-tetrahydrofolate
NAA	*N*-acetyl-l-aspartate
NABQI	*N*-acetyl benzoquinoneimine
NAC	*N*-acetylcysteine
NAD^+/NADH	Nicotinamide adenine dinucleotide (oxidized/reduced)
$NADP^+$	Nicotinamide adenine dinucleotide phosphate
NADPH	Nicotinamide dinucleotide phosphate (reduced)
NAFLD	Nonalcoholic fatty liver disease
NCC	Sodium–chloride co-transporter
ncRNA	Noncoding RNA
NEFA	Nonesterified fatty acid
NeuAc	Neuraminic (sialic) acid
NFAT2	Transcription factor; nuclear factor of activated T cells-2
NFκB	Nuclear factor kappa-light-chain-enhancer of activated B cell
NGF	Nerve growth factor
NGS	Next generation sequencing
NHE	Sodium/hydrogen exchanger
NIDDM	Noninsulin-dependent diabetes mellitus
NK	Natural killer cells
NKCC1	Na^+ K^+ and Cl^- cotransporter
NKCC2	Sodium–potassium–chloride co-transporter
NKH	Nonketotic hyperglycinemia
NLR	NOD-like receptor
NMDA	*N*-methyl-d-aspartate
NMR	Nuclear magnetic resonance
NO	Nitric oxide
NOS	Nitric oxide synthase
NPC1L1	Niemann–Pick C1-like protein
NPY	Neuropeptide Y
nt	Nucleotide
NTX	N-terminal (Amino-terminal) telopeptide

OAA	Oxaloacetate
OGTT	Oral glucose tolerance test
OI	Osteogenesis imperfecta
OMM	Outer mitochondrial membrane
ONDST	Overnight dexamethasone suppression test
OPG	Osteoprotegerin
OSF-1	Osteoblast-stimulating factor 1
o-Tyr	Ortho-tyrosine
P1CP	Procollagen type 1 *C*-terminal peptide
P1NP	Procollagen type 1 *N*-terminal propeptide
p38	Stress-activated protein kinase
p53	Tumor-suppressor protein
p62	Nucleoporin
PA	Phosphatidic acid
PAF	Platelet-activating factor
PAGE	Polyacrylamide gel electrophoresis
PAI-1	Plasminogen activator inhibitor type 1
PAMP	Pathogen-associated molecular pattern
PAPS	Phosphoadenosine-5′-phosphosulfate
PAR2	Protease-activated receptor 2
PBG	Porphobilinogen
PC	Phosphatidylcholine
PC	Pyruvate carboxylase
PCD	Programmed cell death
PCI	Percutaneous coronary intervention
pCO_2	Partial pressure of carbon dioxide
PCP	Phencyclidine
PCR	Polymerase chain reaction
PCSK9	Proprotein convertase subtilisin/kexin type 9
PDE	Phosphodiesterase
PDGF	Platelet-derived growth factor
PDH	Pyruvate dehydrogenase
PDK1	Phosphoinositide-dependent kinase 1
PE	Phosphatidylethanolamine
PECAM-1	Platelet/cell-adhesion molecule 1 (CD31)
PEM	Protein energy malnutrition
PEP	Phosphoenolpyruvate
PEPCK	Phosphoenolpyruvate carboxykinase
PEST	ProGluSerThr degradation signal
PET/MRI	Positron emission tomography/magnetic resonance imaging
PFK	Phosphofructokinase
PFK-2/Fru-2,6-BPase	Phosphofructokinase-2/fructose-2, 6-bisphosphatase
PG	Prostaglandins
PGE_2	Prostaglandin E_2
PGI_2	Prostaglandin I_2
PH	Pleckstrin homology domains
pI	Isoelectric point
PI	Phosphatidylinositol
PI	Propidium iodide
PI3K	Phosphatidylinositol-3-kinase
PIP_2	Phosphatidylinositol 4,5- bisphosphate
PIP_3	Inositol 1, 4,5-bisphosphate
PIP_3	Phosphatidylinositol 3,4,5-trisphosphate
PK	Protein kinase: PKA, PKC
PK	Pyruvate kinase
PKA	Protein kinase A
PKC	Protein kinase C
PKU	Phenylketonuria
PL	Phospholipase: PLA_2, PLC, PLC-β, PLD
PLP	Pyridox(am)ine-5′-phosphate oxidase
PNH	Paroxysmal nocturnal hemoglobinuria
PNPO	Pyridoxal phosphate
PNS	Peripheral nervous system
pO_2	Partial pressure of oxygen
POMC	Proopiomelanocortin
PP2A	Protein phosphatase-2A
PPAR	Peroxisome proliferator–activated receptor
PPi	Pyrophosphate
Prot	Protein
PRPP	Phosphoribosyl pyrophosphate
PRR	Pattern-recognition receptors
PS	Phosphatidylserine
PSA	Prostate-specific antigen
PT	Prothrombin time
PTA	Plasma thromboplastin antecedent
PTEN	Phosphatase and TENsin homologue
PTH	Parathyroid hormone
PTK	Protein tyrosine kinase
PTM	Posttranslational modification
PTPase	Phosphotyrosine phosphatase
PUFA	Polyunsaturated fatty acid
PXR	Pregnane X receptor
Q	Ubiquinone/ubiquinol
RA	Rheumatoid arthritis
RABP	Retinoic acid–binding protein
RAE	Retinol activity equivalent
Raf	Family of serine/threonine kinases
RANK	Receptor activator of nuclear factor NFκB
RANKL	RANK ligand
Rap	Small GTPase
RAR	Retinoid acid receptor
Ras	Small monomeric G-protein; a GTPase
Rb	Retinoblastoma protein
RBC	Red blood cell
RBP	Serum retinol-binding protein
RDA	Recommended dietary allowance
RER	Respiratory exchange rate
RER	Rough endoplasmic reticulum
RFLP	Restriction fragment length polymorphism
RGD	Arg-Gly-Asp recognition sequence
Rheb	Ras homologue enriched in brain
Rho GEF	Rho GTPase guanine nucleotide exchange factor
Rictor	mTORC-2 complex
RIP	Receptor-interacting protein
RIP1	Serine/threonine kinase

RISC	RNA-induced silencing complex
RLR	RIG-1-like receptor
RMR	Resting metabolic rate
RNA	Ribonucleic acid
RNApol	RNA polymerase
RNAseq	Deep sequencing technology for transcriptome analysis
RNS	Reactive nitrogen species
ROS	Reactive oxygen species
ROTEM	Rotational thromboelastometry
RP-HPLC	Reversed phase HPLC
rRNA	Ribosomal RNA
RSK1	A kinase
rT3	Reverse T3
RTA	Renal tubular acidosis
RXR	Retinoid X receptor
S	Substrate
S	Synthetic phase of interphase
S6K1	Ribosomal S6 kinase 1
SAC	Spindle assembly checkpoint
SAM	S-adenosylmethionine
SCAP	SREBP-cleavage-activating protein
SCD	Sickle cell disease
SCD	Stearoyl-CoA desaturase
SCFA	Short-chain fatty acid
SCID	Severe combined immunodeficiency
SCIDS	Severe combined immunodeficiency syndrome
SDS	Sodium dodecylsulfate
SDS-PAGE	Sodium dodecylsulfate-polyacrylamide gel electrophoresis
SECIS	Selenocysteine insertion sequence
SGLT	Na^+-coupled glucose transporter
SGLT-1	Sodium/glucose-linked transport-1 (a membrane transporter)
SGOT	Serum glutamate oxaloacetate transaminase
SGPT	Serum glutamate pyruvate transaminase
SH2	Src-homology region 2
SHBG	Sex hormone–binding globulin
Shc	Src-homology and collagen-like adapter proteins
SHP	SH2 domain-containing phosphatase
SIADH	Syndrome of inappropriate antidiuretic hormone secretion
sIg	Surface Ig
siRNA	Small interfering RNA
SLE	Systemic lupus erythematosus
SMPDB	Small Molecule Pathway Database
SNO-Hb	S-nitrosohemoglobin
snoRNA	Small ribonuclear RNA
snoRNP	Small ribonuclear protein complex
SNP	Single nucleotide polymorphism
SOD	Superoxide dismutase
SOS	Son of Sevenless, guanine nucleotide exchange factor
SPCA	Serum prothrombin conversion accelerator
SpO_2	Peripheral capillary oxygen saturation
SR	Sarcoplasmic reticulum
Src	Homology and collagen domain protein, a nonreceptor PTK
Src	A tyrosine kinase
SRE	Sterol regulatory element
SREBP	Sterol regulatory element-binding protein
SRP	Signal recognition particle
SSB	Sugar-sweetened beverage
STAT1	Signal transducer and activator of transcription-1
STATs	Signal transducer and activators, transcription factors
Succ-CoA	Succinyl-CoA
T1D	Type 1 diabetes mellitus
T2D	Type 2 diabetes mellitus
T3	Triiodothyronine
T4	Thyroxine
TAG, TG	Triacylglycerol, also known as triglyceride
T-ALL	T lymphocyte–acute lymphoblastic leukemia
TBG	Thyroxine-binding globulin
TC-10	A G-protein
TCA	Tricarboxylic acid cycle
TCA cycle	Tricarboxylic acid cycle
TCI	Transcobalamin I
TCII	Transcobalamin II
TCR	T-cell receptor
TCT	Thrombin clotting time
TdT	Terminal deoxynucleotidyl transferase
TEG	Thromboelastography
TF	Transcription factor
T_{FH}	T follicular helper cells
TFIIH	General transcription factor
TFPI	Tissue factor pathway inhibitor
Tg	Thyroglobulin
TG	Triacylglycerol (also triglyceride)
TGF-α	Transforming growth factor-alpha
TGF-β	Transforming growth factor-beta
T_H	T helper cells ($CD4^+$ T cells)
THF	Tetrahydrofolate
THRB	Thyroid hormone receptor beta gene
TIM	Transporter in inner membrane
TIMP	Tissue inhibitor of MMPs
TKI	Tyrosine kinase inhibitor
TLR	Toll-like receptor
T_{max}	Maximum rate of transport
TNF	Tumor necrosis factor
TNFR	Death receptor, TNF family member
TNF-α	Tumor necrosis factor α
TOM	Transporter in outer membrane
tPA	Tissue-type plasminogen activator
TPN	Total parenteral nutrition
TPO	Thyroid peroxidase

TPP	Thiamine pyrophosphate
TRADD	TNF-receptor-associated death domain
TRAFs	TNF-receptor-associated factors
TRAIL	Death receptor, TNF family member
Tregs	T regulatory cells, suppressive T cells
TRH	Thyrotropin-releasing hormone, thyreoliberin
tRNA	Transfer RNA
TSC1/2	Tuberous sclerosis complex 1/2
TSH	Thyroid-stimulating hormone
TSS	Transcriptional start site
T-tubule	Transverse tubule
TWEAK	Death receptor, TNF family member
TX	Thromboxane
TXA_2	Thromboxane A_2
UAS	Upstream activation sequence
UCP	Uncoupling protein
UDP-Gal	Uridine diphosphate galactose
UDP-GalNAc	Uridine diphosphate N-acetylgalactosamine
UDP-Glc	Uridine diphosphate glucose
UDP-GlcNAc	Uridine diphosphate N-acetylglucosamine
UDP-Xyl	Uridine diphosphate xylose
UFC	24-h urine free cortisol
UFH	Unfractionated heparin
UKPDS	UK Prospective Diabetes Study
uPA	Urinary-type plasminogen activator
UPR	Unfolded protein response
UPS	Ubiquitin–proteasome system
URL	Upper reference limit
UTR	Untranslated region
UVRAG	UV radiation resistance–associated gene protein
Va/Q	The ratio of ventilation to perfusion
VAChT	Acetylcholine transporter
VCAM-1	Vascular cell adhesion molecule 1
VDCC	Voltage-dependent Ca^{2+}-channel
VEGF	Vascular endothelial growth factor
VIP	Vasoactive intestinal peptide
VLDL	Very-low-density lipoprotein(s)
V_{max}	Maximum velocity
VO_2	Oxygen consumption rate
VP	Vasopressin; antidiuretic hormone
VSMC	Vascular smooth muscle cell(s)
vWF	Von Willebrand factor
WHO	World Health Organization
Wnt	A signaling pathway related to cell growth and proliferation; the pathway abbreviation relates to "Wingless-related integration site"
XP	Xeroderma pigmentosum
ZAP-70	PTK that is essential for antigen-dependent T-cell activation
ZF	Zona fasciculata
ZG	Zona glomerulosa
ZR	Zona reticularis

CHAPTER

1 Biochemistry and Clinical Medicine: Introduction and Overview

*John W. Baynes and Marek H. Dominiczak**

INTRODUCTION

Biochemistry is constantly changing

This has been a mantra of our book since its conception. A new edition is a welcome opportunity to review the progress in the field.

Research into the human **genome** and particularly gene regulation continues to drive medical progress together with three other expanding fields: the study of the **transcriptome**, the **proteome**, and the **metabolome** (Fig. 1.1). New data fill in the gaps, completing the knowledge about receptors, signaling pathways, transcription factors, and bidirectional information transfer between each of these large domains. From the biochemistry perspective, perhaps the most exciting development in the last few years has been the expansion of knowledge of proteins participating in the transfer of external metabolic signals to intracellular pathways and to and from the genome and the role of these networks in the regulation of cell division and growth (Fig. 1.2). This has also provided new insights into the pathogenesis of cancer and chronic disease and enabled development of new therapies for a range of diseases, including novel vaccines for protection against SARS-CoV-2, the virus that causes COVID-19, and emergent technologies for editing the genome and altering gene expression.

All this has changed the way we look at metabolism. In addition to the strings of chemical reactions that have been the essence of biochemistry since its inception, we now recognize cascades of signaling molecules and receptors that involve hormones and neurotransmitters in both direct control of reactions and in the control of synthesis of enzyme proteins. These developments open up enormous avenues for therapeutic drug development but also pose challenges for the student of biochemistry, such as sometimes complex protein terminology. The need to become familiar with often nonintuitive abbreviations and acronyms that identify signaling molecules and transcription factors is now a fact of life in biochemistry.

Another recent development that changes our perspective on biochemistry is the impact of structural biology: better understanding of phenomena, such as membrane trafficking, and the nature of cellular scaffolds that house chemical processes. We learn that physical structures compartmentalize and modify many processes, and that they play a fundamental role in a developing disease.

Diseases that are related to nutrition, lifestyle, and environment, obesity, diabetes, atherosclerosis, and cardiovascular disease, on the one hand, and malnutrition and nutritional deficiencies, on the other, continue to be major global health concerns. Physicians spend an increasing fraction of their time dealing with the prevention and treatment of these chronic and age-related diseases.

Our expanding computing ability and use of big data from combined clinical and population studies allow the development of algorithms that predict the risk of disease, departing from a binary system of risk assessment to a continuous one, where individual risk can be more precisely graded.

Finally, as molecular neuroscience keeps developing, we appreciate better the enormous complexity of the brain, at the same time becoming aware that on the molecular level many neurophysiological and pathological processes can now be understood in terms of alterations in neuronal proteins, signaling cascades, and ion flows.

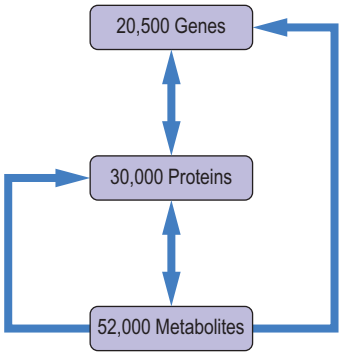


Fig. 1.1 **Human genes, proteins, and metabolites.** Data based on Human Genome Project, Human Proteome Map, and Human Metabolome Database (see websites). Numbers are approximate.

*The authors gratefully acknowledge the original contributions to this chapter by Dr. Naoyuki Taniguchi, Department of Glyco-Oncology and Medical Biochemistry, Osaka International Cancer Institute, Japan.

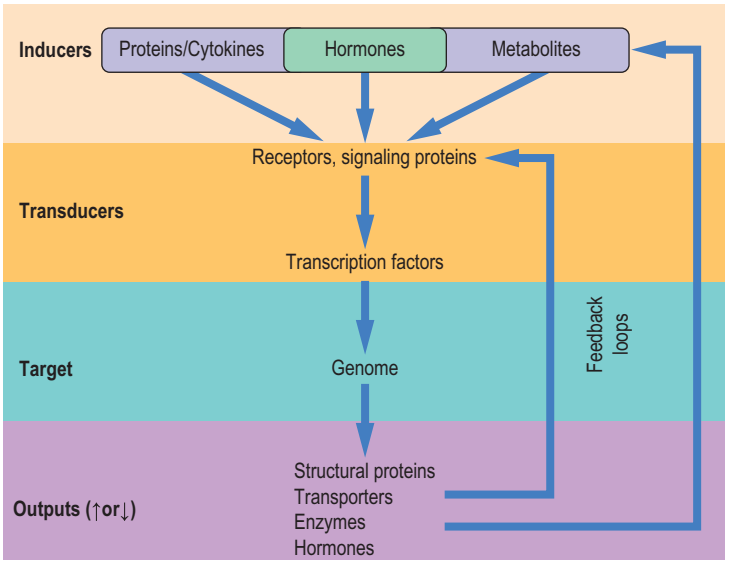


Fig. 1.2 **Integrated view of regulatory loops between signaling proteins, enzymes, genome, and metabolism.** Note that many signaling proteins are also enzymes. There are both protein and nonprotein (metabolite-derived) hormones.

WHAT IS THE SIXTH EDITION OF MEDICAL BIOCHEMISTRY?

Biochemistry has fuzzy borders

Biochemistry is not a discipline with clear borders. It links seamlessly to fields such as cell biology, anatomy, physiology, and pathology. In fact, it is not possible to solve a clinical problem without crossing interdisciplinary borders. In this book, we cross these borders frequently, both in the main text and in clinical boxes. Thus the chapters that cover **nutrition, water and electrolytes, acid–base balance**, and **specialized tissues** are both interdisciplinary and integrative with regard to previously presented aspects of biochemistry. From its first edition this book put more emphasis on water, electrolytes, and acid–base balance than many other textbooks. We continue to insist on the importance of these topics, particularly for the student of medicine, and we believe these should be taught as a part of biochemistry courses, providing a foundation for clinical chemistry and medical practice.

We wrote *Medical Biochemistry* because we have been convinced that understanding biochemistry improves the practice of medicine. It is essential for a physician to integrate new developments into everyday practice. What only a few years ago was theory and speculation is now a part of the tool kit used during ward rounds and in case conferences. The question we asked ourselves many times during the writing process was, "How could this piece of information improve your clinical reasoning?" Therefore, throughout the text, we highlight issues that a physician encounters at the bedside and link them to basic concepts.

We believe that a textbook should not only provide a base of knowledge essential for a doctor, but it should also flag emerging topics that are likely to be incorporated into practice in the foreseeable future, and perhaps instill in the student the need for awareness of the unexpected.

The real world confirms that this is essential. The catastrophic COVID-19 pandemic highlighted the importance of robust public health systems at national and international levels and the significance of having flexible, research-based diagnostic laboratories that can quickly cope with unexpected challenges, but above all, the societal importance of cutting-edge science that can come to the rescue at times of crisis.

Improvements in the sixth edition

Medical Biochemistry is not designed to be a review text or resource for preparation for multiple-choice exams. It is a resource for your clinical career. Our text is shorter than many in our discipline, and it focuses on explanation of key concepts and their integration into the concept of disease entities. We hope that this approach will help your future clinical practice.

During the preparation of the sixth edition, as before, we sought to update the facts and improve the quality of explanation. We maintained the structure of the previous edition. We start with **Molecules and Cells** and move on to **Metabolism**, the core of biochemistry. Then we address the **Molecular Basis of Inheritance**, discussing **deoxyribonucleic acid (DNA), ribonucleic acid (RNA)**, protein synthesis, regulation of gene expression, and the systemic approaches: genomics, proteomics, and metabolomics. We also address **Cellular Signaling and Growth** and their links to aging and cancer. We then proceed to the "integrating" subjects. In **Fuels and Nutrients**, we discuss the gastrointestinal tract and the diseases associated with

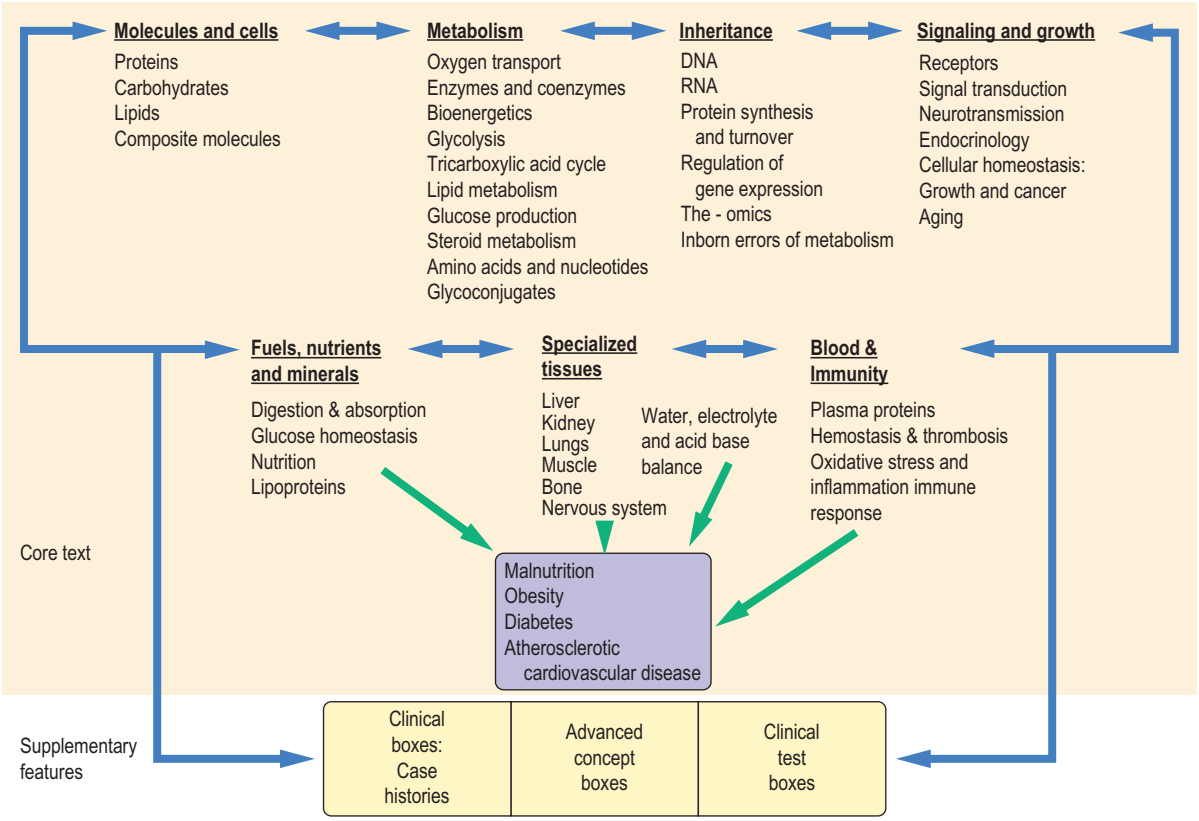


Fig. 1.3 ***Medical Biochemistry*, sixth edition: Map of the Book.** All sections of the book are strongly interrelated. As you learn about metabolism, you'll gain the knowledge of the key inherited metabolic errors and their effects. Major contemporary health issues, such as diabetes mellitus, atherosclerosis, obesity, and malnutrition, are emphasized in relevant chapters. Throughout the book, you will find Clinical Boxes and Clinical Test Boxes, which integrate basic science with clinical practice, and Advanced Concept Boxes that expand on selected issues. Updated literature and web references will facilitate further study.

glucose and lipoprotein metabolism, which are currently major public health issues. This is continued in **Specialized Tissues and Their Function** section, where we describe the function of the liver, muscle, and the brain. Here we also discuss metabolism of foreign substances, including therapeutic drugs. The last section, **Blood and Immunity**, describes the body defense mechanisms: hemostasis (blood coagulation) and the immune response, the role of inflammation, and the consequences of oxidative stress. Fig. 1.3 shows our Map of the Book and illustrates the integration of biochemical topics in a broader biological context. Fig. 1.4 illustrates how you can learn also by starting your study with an integrating chapter and then fill in the details by going to chapters discussing a particular topic in greater depth in a form of "reverse learning."

The references and websites listed at the end of the chapters expand on what has been presented, should you wish to look for more details.

In this edition, there are many more links to videos that present complex biochemical processes in a dynamic fashion. Frankly, some processes are difficult to capture in static text and figure format. Further, through its **Student Consult** program, Elsevier provides links to expanded discussions in textbooks in anatomy, cell biology, microbiology, physiology, pharmacology, immunology, pathology, and clinical chemistry. These resources are conveniently hyperlinked to topics in biochemistry.

WHAT IS BIOCHEMISTRY: THE SHORT OVERVIEW

Studying biochemistry helps to understand the interplay of nutrition, metabolism, and genetics in health and disease.

The human organism is, on the one hand, a tightly controlled, integrated, and self-contained metabolic system. Claude Bernard (1813–78), a French physiologist, developed the concept of the internal environment (*milieu intérieur*). On the other hand, it is an open system that communicates with its environment. Despite these two seemingly contradictory characteristics, the body manages to maintain its internal environment for decades.

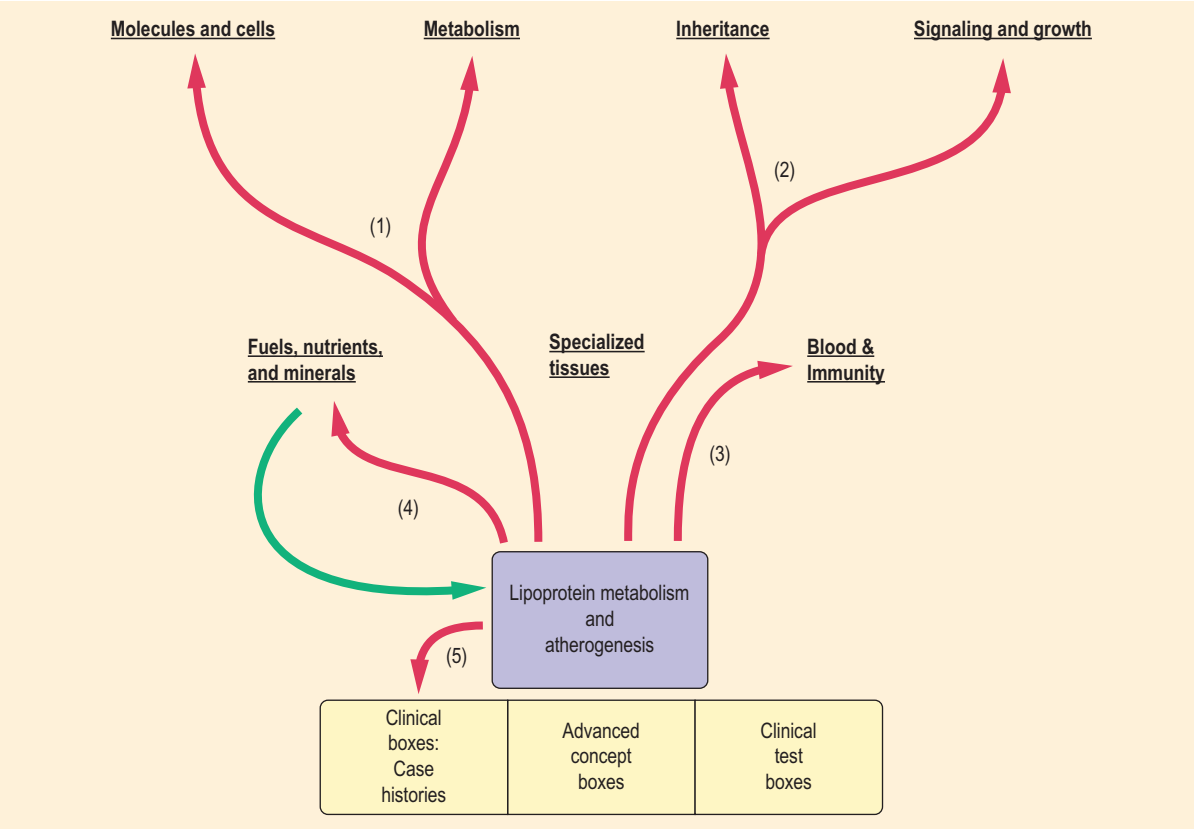


Fig. 1.4 **The reverse learning method: Exploring a topic in *Medical Biochemistry*.** The chapters of obesity, diabetes, and lipids-atherosclerosis integrate many fields of knowledge. One can proceed from such integrated topics to the other parts of the book to deepen one's knowledge. Let's assume that one reads the chapter on lipoprotein metabolism and atherogenesis (Chapter 33) in the section on Fuels Nutrients and Minerals (green arrow). The red arrows point to different parts of the book when one can find more detailed information on structure of the compounds involved (1), specific mechanisms that underlie vascular remodeling (2), processes underlying cellular accumulation in the atherosclerotic plaque (3), and aspects of dietary prevention (4). Any clinical topic will be reinforced by looking up the clinical boxes (5).

We regularly top up on fuel (consume food) and water and take up oxygen from inspired air to use for oxidative metabolism, a highly regulated sequence of low-temperature combustion reactions. We then use the energy generated from metabolism to perform work and maintain body temperature. We get rid of (exhale or excrete) carbon dioxide, water, and nitrogenous waste. The amount and quality of food we consume have a significant impact on our health: malnutrition, on the one hand, and obesity and diabetes, on the other, are currently major public health issues worldwide.

Proteins, carbohydrates, and lipids are the major structural components of the body

Proteins are building blocks and catalysts; as structural units, they form the "architectural" framework of tissues; as enzymes, together with helper molecules (**coenzymes** and **cofactors**), they catalyze biochemical reactions. Proteins also play a fundamental role in the transfer of information (**signaling**) at both the cell and the entire-organism level, processes that are essential for the function of DNA and the regulation of **gene expression**.

Carbohydrates and **lipids** as monomers or relatively simple polymers are our major **energy sources**. They can be stored in tissues as glycogen and triglycerides. Lipids such as **cholesterol** and **phospholipids** form the backbone of biological membranes. Carbohydrates and lipids can also be linked to both proteins and lipids, forming complex structures (glycoconjugates) essential for **cell-signaling systems** and processes such as cell **adhesion** and **immunity**.

Chemical variables, such as **pH, oxygen tension**, and **inorganic ion and buffer concentrations**, define the homeostatic environment in which metabolism takes place. Minute changes in this environment - for example, a change in just a few degrees in body temperature or a tenth of a pH unit - can be life-threatening.

Biological membranes partition metabolic pathways into different cellular compartments. These water-impermeable structures are dotted with an array of "doors and gates"

(membrane transporters) and "locks" that accept a variety of keys, including hormones and cytokines that initiate intracellular signaling cascades. Membranes play a fundamental role in **ion** and **metabolite transport** and in **signal transduction**, both within individual cells and between cells. In fact, most of the body's energy is consumed to generate heat and to maintain ion and metabolite gradients across membranes. Cells throughout the body are critically dependent on **electrical and chemical potentials across membranes** for nerve transmission, muscle contraction, nutrient transport, and the maintenance of cell volume.

Carbohydrates and lipids are our primary sources of energy, but our nutritional requirements also include **amino acids** (components of proteins), **inorganic molecules** (sodium, calcium, potassium, chloride, bicarbonate, phosphate, and others), and micronutrients - **vitamins** and **trace elements**.

Glucose (present in blood in a pure form and stored in the form of glycogen) is metabolized through **glycolysis**, a universal non-oxygen-requiring (anaerobic) pathway for energy production. It yields **pyruvate**, setting the stage for oxidative metabolism in the mitochondrion. It also generates metabolites that are the starting points for the synthesis of **amino acids, proteins, lipids**, and **nucleic acids**.

Glucose is the most important fuel for the brain; therefore maintaining its concentration in plasma is essential for survival. Glucose homeostasis is regulated by the hormones that coordinate metabolic activities among cells and organs - primarily insulin and glucagon but also epinephrine and cortisol.

Oxygen is essential for energy production but can also be toxic

During aerobic metabolism, pyruvate, the end product of anaerobic glycolysis, is transformed into **acetyl coenzyme A (acetyl-CoA)**, the common intermediate in the metabolism of carbohydrates, lipids, and amino acids. Acetyl-CoA enters the central metabolic engine of the cell, the **tricarboxylic acid (TCA) cycle** located in mitochondria. Acetyl-CoA is oxidized to **carbon dioxide** and reduces the important coenzymes **nicotinamide adenine dinucleotide (NAD^+)** and **flavin adenine dinucleotide (FAD)**. Reduction of these nucleotides captures the energy from fuel oxidation.

Most of the energy in biological systems is obtained by **oxidative phosphorylation**. This process involves oxygen consumption, or **respiration**, by which the organism oxidizes NADH and $FADH_2$ in the mitochondrial **electron transport** chain (ETC) to produce a hydrogen ion gradient across the inner mitochondrial membrane. The energy in this **electrochemical gradient** is then transformed into the chemical energy of **adenosine triphosphate (ATP)**. Biochemists call ATP the "common currency of metabolism" because it allows energy produced by fuel metabolism to be used for work, transport, and biosynthesis. Although oxygen is essential for aerobic metabolism, it can also cause **oxidative stress** and widespread tissue damage during **inflammation**. Powerful **antioxidant defenses** exist to protect cells and tissues from the damaging effects of reactive oxygen.

Metabolism continuously cycles the eating, postprandial, and fasting states

The direction of the main pathways of carbohydrate and lipid metabolism changes in response to food intake. In the fed state, the active pathways are **glycolysis, glycogen synthesis, lipogenesis**, and **protein synthesis**, rejuvenating tissues and storing the excess of metabolic fuel. In the fasting state, the direction of metabolism reverses: glycogen and lipid stores are degraded through **glycogenolysis** and **lipolysis**, providing a constant stream of **substrates for energy production**. As glycogen stores become depleted, proteins are degraded to produce carbon skeletons for synthesis of glucose through **gluconeogenesis**, guaranteeing a constant supply of blood sugar, while other biosynthetic pathways are slowed down. Poor diets and alterations in fuel transport and metabolism underlie chronic and ubiquitous pathologies, such as **diabetes mellitus, obesity**, and **atherosclerotic cardiovascular disease**, which are currently major public health issues.

Tissues perform specialized functions

Specialized tissue functions include muscle contraction; nerve conduction; bone formation; immune surveillance; hormonal signaling; maintenance of pH, fluid, and electrolyte balance; and detoxification of foreign substances. **Glycoconjugates** (glycoproteins, glycolipids, and proteoglycans) are needed for tissue organization, structural integrity, and cell-to-cell communications. Recent progress in understanding cellular signaling systems has improved our insight into **cell growth** and **repair mechanisms**. Their time-dependent decline leads to **aging**, and their failure causes age-related diseases, including **neurodegenerative diseases** and **cancer**.

The genome underpins it all

The genome provides the mechanism for conservation and transfer of genetic information through regulation of the expression of constituent genes and control of protein synthesis. The synthesis of individual proteins is controlled by information encoded in DNA and transcribed into RNA, which is then translated into peptides that fold into **functional protein molecules**. The spectrum of expressed proteins and the control of their temporal expression during development, adaptation, and aging are responsible for our protein makeup, our **proteome. Epigenetics**, a fast-expanding field, is the study of gene regulation caused by modification of DNA function by means other than changes in its nucleotide sequence. It provides deeper insight into the regulation of gene expression.

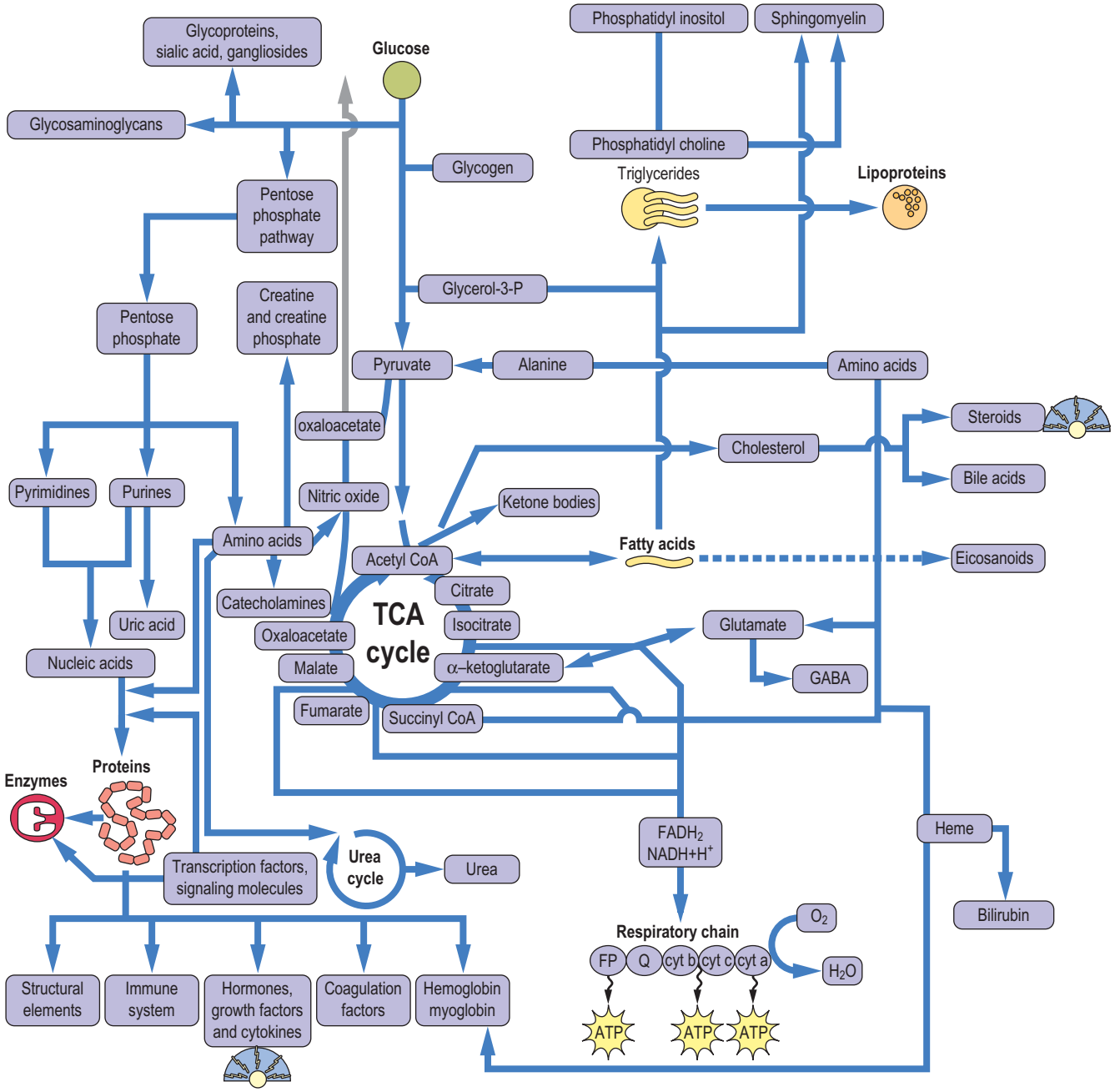


Fig. 1.5 **Biochemistry: All in one.** Here is a bird's-eye view of the field, focusing on metabolism and bioenergetics. It may help structure your study or revision. Refer to it as you study the following chapters, and see how your perspective on biochemistry broadens. ATP, adenosine triphosphate; cyt, cytochrome; FAD/$FADH_2$, flavin adenine dinucleotide (oxidized/reduced); FP, flavoprotein; GABA, γ-aminobutyric acid; glycerol-3-P, glycerol-3-phosphate; NAD^+/NADH, nicotinamide adenine dinucleotide (oxidized/reduced); Q, ubiquinone/ubiquinol; TCA cycle, tricarboxylic acid cycle.

In the past decade, **bioinformatics**, genome-wide association studies, and epigenetics have provided truly fascinating insights into the complexity of genetic regulatory networks. Further, applications of **recombinant DNA** technology have revolutionized the work of clinical laboratories and have recently provided novel tools for editing the genome. The ability to scan the entire genome and the information obtained by **genomics, proteomics**, and **metabolomics** provide new insights into gene regulation, protein synthesis, and metabolism.

All of this is summarized in Fig. 1.5, a complex scheme that resembles the plan of the London Tube (Further Reading). Like the Tube, with its many stations, biochemistry is navigable with a good map; don't be intimidated by the many as-yet unfamiliar terms. Refer to this figure as you progress in your studies, and you will notice how your understanding of biochemistry improves.

FURTHER READING

Atkins P. *What is chemistry?* Oxford, UK: Oxford University Press; 2013.

Cooke M, Irby DM, Sullivan W, et al. American medical education 100 years after the Flexner Report. *N Engl J Med.* 2006;355:1339–1344.

Dominiczak MH. Teaching and training laboratory professionals for the 21st century. *Clin Chem Lab Med.* 1998;36:133–136.

Dominiczak, M. H. (2012). Contribution of biochemistry to medicine: Medical biochemistry and clinical biochemistry. *UNESCO encyclopedia of life support systems (UNESCO-EOLSS)*. http:// www.eolss.net.

Ludmerer KM. Learner-centered medical education. *New England Journal of Medicine.* 2004;351:1163–1164.

Powell D, Carraccio C. Toward competency-based medical education. *N Engl J Med.* 2018;378:3–5.

RELEVANT WEBSITES

Human Metabolome Database (HMDB), version 3.6: http://www.hmdb.ca/

Human Proteome Map: http://www.humanproteomemap.org/

\

Overview of the Human Genome Project, NIH National Human Genome Research Institute: https://www.genome.gov/12011238/an-overview-of-the-human-genome-project/

\

Transport for London. (n.d.). Tube map. Retrieved from http://www.tfl.gov.uk/assets/downloads/standard-tube-map.pdf

ABBREVIATIONS

Acetyl-CoA	Acetyl coenzyme A
ATP	Adenosine triphosphate
COVID-19	Coronavirus disease 2019. Infectious disease caused by SARS-CoV-2
Cyt a, b, c	Cytochrome a, cytochrome b, cytochrome c
DNA	Deoxyribonucleic acid
$FAD/FADH_2$	Flavin adenine dinucleotide (oxidized/reduced)
FP	Flavoprotein
GABA	γ-Aminobutyric acid
Glycerol-3-P	Glycerol-3-phosphate
$NAD^+/NADH$	Nicotinamide adenine dinucleotide (oxidized/reduced)
Q	Ubiquinone/ubiquinol
RNA	Ribonucleic acid
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2. The virus that causes COVID-19
TCA cycle	Tricarboxylic acid cycle

This page intentionally left blank

CHAPTER
2 Amino Acids and Proteins

Ryoji Nagai

LEARNING OBJECTIVES

After reading this chapter, you should be able to:

- Classify the amino acids based on their chemical structure and charge.
- Explain the meaning of the terms pK_a and pI as they apply to amino acids and proteins.
- Describe the elements of the primary, secondary, tertiary, and quaternary structure of proteins.
- Describe the principles of ion-exchange, gel filtration, and affinity chromatography, and electrophoresis and isoelectric focusing, and describe their application in protein isolation and characterization.

INTRODUCTION

Proteins are major structural and functional polymers in living systems

Proteins have a broad range of activities, including catalysis of metabolic reactions and transport of vitamins, minerals, oxygen, and fuels. Some proteins make up the structure of tissues; others function in nerve transmission, muscle contraction, and cell motility, with still others in blood clotting and immunologic defenses and as hormones and regulatory molecules. Proteins are synthesized as a sequence of amino acids linked together in a linear polyamide (polypeptide) structure, but they assume complex three-dimensional shapes in performing their functions. There are about 300 amino acids present in various animal, plant, and microbial systems, but **only 20 amino acids are coded by DNA to appear in proteins**. Many proteins also contain modified amino acids and accessory components, termed prosthetic groups, such as carbohydrates, lipids, and more complex structures, such as heme in hemoglobin (Chapter 5). A range of chemical techniques are used to isolate and characterize proteins by a variety of criteria, including mass, charge, and three-dimensional structure. Proteomics is an emerging field that studies the full range of expression of proteins in a cell or organism and changes in protein expression in response to growth, hormones, stress, and aging.

ADVANCED CONCEPT BOX
NONPROTEIN AMINO ACIDS

Measurement of abnormal amino acids or elevated concentrations of amino acids in urine (aminoaciduria) is useful for clinical diagnosis (Chapter 15). In plasma, free amino acids are usually found at 10–100 μmol/L, including many that are not found in protein. Citrulline, for example, is a metabolite of L-arginine and a product of nitric oxide synthase, an enzyme that produces nitric oxide, an important vasoactive signaling molecule. Creatinine is an amino acid derived largely from muscle and is excreted in relatively constant amounts per unit body mass per day. Thus the creatinine concentration in urine, normally about 1 mg/mL, can be used to correct for urine dilution. Dipeptides such as carnosine, β-alanyl-L-histidine, and anserine, β-alanyl-*N*-methylhistidine, can be detected in the skeletal muscle and brain, and those dipeptides act as scavengers for reactive oxygen species (Chapter 42).

AMINO ACIDS

Amino acids are the building blocks of proteins

Stereochemistry: Configuration at the α-carbon and D- and L-isomers

Each amino acid has a central carbon, called the α-carbon, to which four different groups are attached (Fig. 2.1):

- A basic amino group ($-NH_2$)
- An acidic carboxyl group (–COOH)
- A hydrogen atom (–H)
- A distinctive side chain (–R)

One of the 20 amino acids, proline, is not an α-amino acid but an α-imino acid. Except for glycine, all amino acids contain at least one asymmetric carbon atom (the α-carbon atom), giving two isomers that are optically active (i.e., they can rotate plane-polarized light). These isomers, referred to as **stereoisomers** or enantiomers, are said to be chiral, a word derived from the Greek word meaning "hand." Such isomers are nonsuperimposable mirror images and are analogous to left and right hands, as shown in Fig. 2.2.

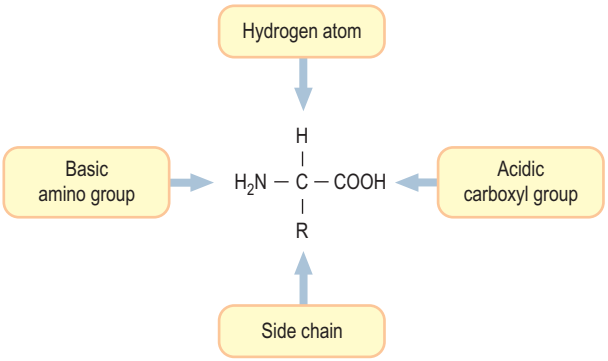


Fig. 2.1 **Structure of an amino acid.** Except for glycine, four different groups are attached to the α-carbon of an amino acid. Table 2.1 lists the structures of the R groups.

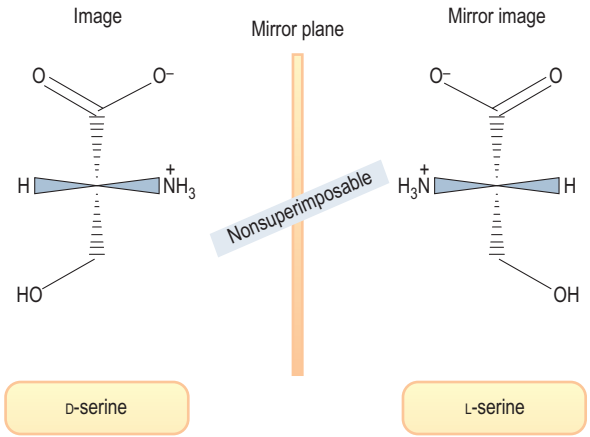


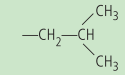
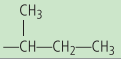
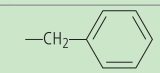
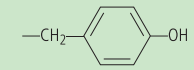
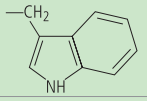
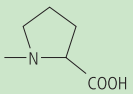
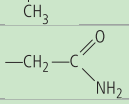
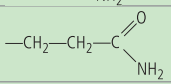
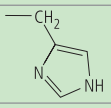
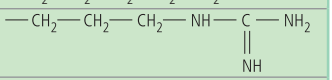
Fig. 2.2 **Enantiomers.** The mirror-image pair of amino acids. Each amino acid represents nonsuperimposable mirror images. The mirror-image stereoisomers are called enantiomers. Only the L-enantiomers are found in proteins.

The two amino acid configurations are called D (for *dextro*, meaning "right") and L (for *levo*, meaning "left"). **All amino acids in proteins are of the L-configuration** because proteins are biosynthesized by enzymes that insert only L-amino acids into the peptide chains.

Classification of amino acids based on chemical structure of their side chains

The properties of each amino acid are dependent on its side chain (–R), which determines the structure and function of proteins and the electrical charge of the molecule. Knowledge of the properties of these side chains is important for understanding methods of analysis, purification, and identification of proteins. Amino acids with charged, polar, or hydrophilic side chains are usually exposed on the surface of proteins. The nonpolar hydrophobic residues are usually buried in the hydrophobic interior, or core, of a protein and are out of contact with water. The 20 amino acids in proteins encoded by DNA are listed in Table 2.1 and are classified according to their side chain functional groups.

Table 2.1 20 amino acids found in proteins*

Amino acids	Structure of R moiety
Aliphatic amino acids	
glycine (Gly, **G**)	–H
alanine (Ala, **A**)	$-CH_3$
valine (Val, **V**)	
leucine (Leu, **L**)	
isoleucine (Ile, **I**)	
Sulfur-containing amino acids	
cysteine (Cys, **C**)	$-CH_2-SH$
methionine (Met, **M**)	$-CH_2-CH_2-S-CH_3$
Aromatic amino acids	
phenylalanine (Phe, **F**)	
tyrosine (Tyr, **Y**)	
tryptophan (Trp, **W**)	
Imino acid	
proline (Pro, **P**)	
Neutral amino acids	
serine (Ser, **S**)	$-CH_2-OH$
threonine (Thr, **T**)	
asparagine (Asn, **N**)	
glutamine (Gln, **Q**)	
Acidic amino acids	
aspartic acid (Asp, **D**)	$-CH_2-COOH$
glutamic acid (Glu, **E**)	$-CH_2-CH_2-COOH$
Basic amino acids	
histidine (His, **H**)	
lysine (Lys, **K**)	$-CH_2-CH_2-CH_2-CH_2-NH_2$
arginine (Arg, **R**)	

**The three-letter and single-letter abbreviations in common use are given in parentheses.*

Aliphatic amino acids

The aliphatic amino acids (alanine, valine, leucine, and isoleucine) have saturated hydrocarbons as side chains. Glycine, which has only a hydrogen side chain, is also included in this group. Alanine has a relatively simple structure, a side chain methyl group, whereas valine, leucine, and isoleucine have isopropyl and sec- and iso-butyl groups. All of these amino acids are hydrophobic.

Aromatic amino acids

Phenylalanine, tyrosine, and tryptophan have aromatic side chains

The nonpolar aliphatic and aromatic amino acids are normally buried in the protein core and are involved in hydrophobic interactions with one another. Tyrosine has a weakly acidic hydroxyl group and may be located on the surface of proteins. Reversible phosphorylation of the hydroxyl group of tyrosine in some enzymes is important in the regulation of metabolic pathways. **The aromatic amino acids are responsible for the ultraviolet absorption of most proteins, which have absorption maxima ~280 nm**. Tryptophan has a greater absorption in this region than phenylalanine or tyrosine. The molar absorption coefficient of a protein is useful in determining the concentration of a protein in solution, based on spectrophotometry. Typical absorption spectra of aromatic amino acids and a protein are shown in Fig. 2.3.

Neutral polar amino acids

Neutral polar amino acids contain hydroxyl or amide side chain groups. Serine and threonine contain side-chain hydroxyl groups. These amino acids are sometimes found at the active sites of catalytic proteins, enzymes (Chapter 6). Reversible phosphorylation of peripheral serine and threonine residues of enzymes is also involved in regulation of energy metabolism and fuel storage in the body (Chapter 12). **Asparagine and glutamine have amide-bearing side chains. These are polar but uncharged under physiologic conditions**. Serine, threonine, and asparagine are the primary sites of linkage of sugars to proteins, forming glycoproteins (Chapter 17).

Acidic amino acids

Aspartic and glutamic acids contain carboxylic acids on their side chains; they are ionized at pH 7 and carry negative charges on their β- and γ-carboxyl groups, respectively. In the ionized state, these amino acids are referred to as aspartate and glutamate, respectively.

Basic amino acids

The side chains of the basic amino acids, lysine and arginine, are fully protonated at neutral pH and therefore positively charged. Lysine contains a primary amino group (NH_2) attached to the terminal ε-carbon of the side chain. The ε-amino group of lysine has a $pK_a \approx 11$. Arginine is the most basic amino acid ($pK_a \approx 13$), and its **guanidine** group exists as a protonated guanidinium ion at pH 7.

Histidine ($pK_a \approx 6$) has an **imidazole** ring as the side chain and functions as a general acid–base catalyst in many enzymes. The protonated form of imidazole is called an imidazolium ion.

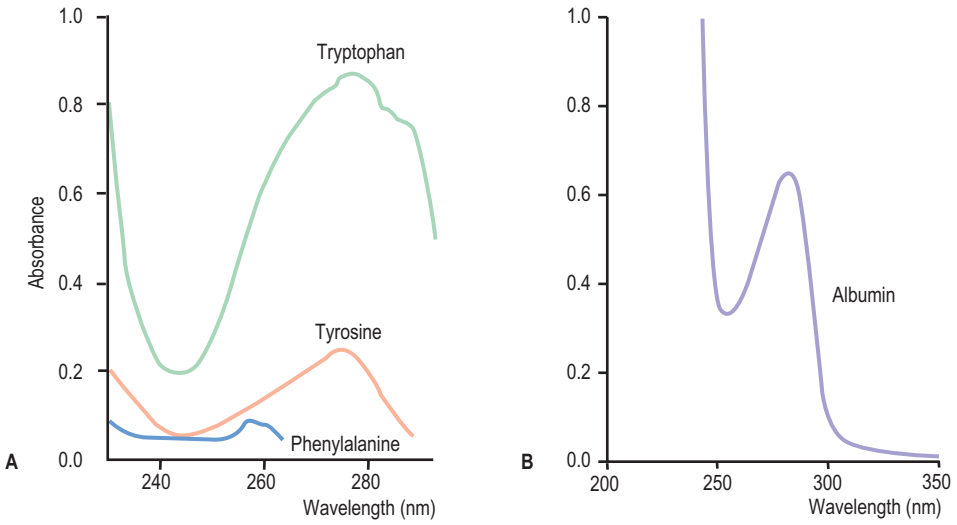


Fig. 2.3 **Ultraviolet absorption spectra of the aromatic amino acids and bovine serum albumin.** (A) Aromatic amino acids such as tryptophan, tyrosine, and phenylalanine have absorbance maxima at 260–280 nm. Each purified protein has a distinct molecular absorption coefficient at around 280 nm, depending on its content of aromatic amino acids. (B) A bovine serum albumin solution (1 mg dissolved in 1 mL of water) has an absorbance of 0.67 at 280 nm using a 1-cm cuvette. The absorption coefficient of proteins is often expressed as $E_{1\%}$ (10 mg/mL solution). For albumin, $E_{1\%}280_{nm} = 6.7$. Although proteins vary in their Trp, Tyr, and Phe content, measurements of absorbance at 280 nm are useful for estimating protein concentration in solutions.

Table 2.2 Summary of the functional groups of amino acids and their polarity

Amino acids	Functional group	Hydrophilic (polar) or hydrophobic (nonpolar)	Examples
acidic	carboxyl, –COOH	polar	Asp, Glu
basic	amine, $–NH_2$	polar	Lys
	imidazole	polar	His
	guanidino	polar	Arg
neutral	glycine, –H	nonpolar	Gly
	amides, $–CONH_2$	polar	Asn, Gln
	hydroxyl, –OH	polar	Ser, Thr
	sulfhydryl, –SH	nonpolar	Cys
aliphatic	hydrocarbon	nonpolar	Ala, Val, Leu, Ile, Met, Pro
aromatic	C-rings	nonpolar	Phe, Trp, Tyr

Sulfur-containing amino acids

Cysteine and its oxidized form, cystine, are sulfur-containing amino acids characterized by low polarity. Cysteine plays an important role in the stabilization of protein structure because it can form a disulfide bond with other cysteine residues, crosslinking protein chains and stabilizing protein structure. Two regions of a single polypeptide chain, remote from each other in the sequence, may be covalently linked through a disulfide bond (intrachain disulfide bond). **Disulfide bonds** are also formed between two polypeptide chains (interchain disulfide bond), forming covalent protein dimers. These bonds can be reduced by enzymes or reducing agents, such as 2-mercaptoethanol or dithiothreitol, to form cysteine residues. Methionine is the third sulfur-containing amino acid and contains a nonpolar methyl thioether group in its side chain.

ADVANCED CONCEPT BOX
LENS DISLOCATION IN HOMOCYSTINURIA (INCIDENCE: 1 IN 200,000)

The most common ocular manifestation of homocystinuria, a defect in sulfur amino acid metabolism (Chapter 15), is lens dislocation occurring around age 10 years. Fibrillin, found in the fibers that support the lens, is rich in cysteine residues. Disulfide bonds between these residues are required for the crosslinking and stabilization of protein and lens structure. Homocysteine, a metabolic intermediate and homolog of cysteine, can disrupt these bonds by homocysteine-dependent disulfide exchange. Another equally rare sulfur amino acid disorder, sulfite oxidase deficiency, is also associated with lens dislocation by a similar mechanism (usually presenting at birth with early refractory convulsions).

Proline, a cyclic imino acid

Proline is different from other amino acids in that it is an α-imino acid with a **pyrrolidine ring** that includes both the α-amino group and the α-carbon. This imino acid forces a "bend" in a polypeptide chain, sometimes causing abrupt changes in the direction of the chain.

Classification of amino acids based on the polarity of the amino acid side chains

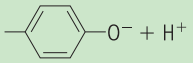
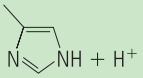
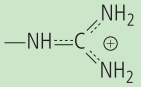
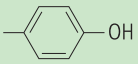
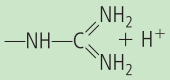
Table 2.2 depicts the functional groups of amino acids and their polarity (hydrophilicity). Polar side chains can be involved in hydrogen bonding to water and to other polar groups and are usually located on the surface of the protein. Hydrophobic side chains contribute to protein folding by hydrophobic interactions and are located primarily in the core of the protein or on surfaces involved in interactions with other proteins.

Ionization state of an amino acid

Amino acids are amphoteric molecules they have both basic and acidic groups

Monoamino and monocarboxylic acids are ionized in different ways in solution, depending on the solution's pH. At pH 7, the "zwitterion" $^{+}H_3N–CH_2–COO^{-}$ is the dominant species of glycine in solution, and the overall molecule is therefore electrically neutral. On titration to acidic pH, the α-amino and carboxyl groups are protonated, yielding the cation

Table 2.3 pK_a values for ionizable groups in proteins

Group	Acid (protonated form) (conjugate acid)	H⁺ + base (unprotonated form) (conjugate base)	pK_a
terminal carboxyl residue (α-carboxyl)	–COOH (carboxylic acid)	$-COO^- + H^+$ (carboxylate)	3.0–5.5
aspartic acid (β-carboxyl)	–COOH	$-COO^- + H^+$	3.9
glutamic acid (γ-carboxyl)	–COOH	$-COO^- + H^+$	4.3
histidine (imidazole)	 (imidazolium)	 (phenolate)	6.0
terminal amino (α-amino)	$-NH_3^+$ (ammonium)	$-NH^+ + H^+$ (amine)	8.0
cysteine (sulfhydryl)	–SH (thiol)	$-S^- + H^+$ (thiolate)	8.3
tyrosine (phenolic hydroxyl)	 (imidazole)	 (guanidinium)	10.1
lysine (ε-amino)	$-NH_3^+$	$-NH_2 + H^+$	10.5
arginine (guanidino)	 (phenol)	 (guanidino)	12.5

Actual pK_a values may vary by several pH units, depending on temperature, buffer, ligand binding, and especially neighboring functional groups in the protein.

$^+H_3N–CH_2–COOH$, whereas titration with alkali yields the anionic $H_2N–CH_2–COO^-$ species:

$$^+H_3N\,CH_2\,COOH \underset{}{\overset{H^+}{\rightleftharpoons}} {}^+H_3N\,CH_2\,COO^- \overset{OH^-}{\rightleftharpoons} H_2N\,CH_2\,COO^-$$

Zwitterion

pK_a values for the α-amino and α-carboxyl groups and side chains of acidic and basic amino acids are shown in Table 2.3. The overall charge on a protein depends on the contribution from basic (positive charge) and acidic (negative charge) amino acid side chains, but the actual charge on the protein varies with the pH of the solution. To understand how the side chains affect the charge on proteins, it is worth recalling the Henderson–Hasselbalch equation.

Henderson–Hasselbalch equation and pK_a

The Henderson-Hasselbalch equation describes the titration of an amino acid and can be used to predict the net charge and isoelectric point of a protein

The general dissociation of a weak acid, such as a carboxylic acid, is given by the equation:

$$HA \rightleftarrows H^+ + A^- \qquad (1)$$

where HA is the protonated form (conjugate acid, or associated form), and A^- is the unprotonated form (conjugate base, or dissociated form).

The dissociation constant (K_a) of a weak acid is defined as the equilibrium constant for the dissociation reaction (1) of the acid:

$$K_a = \frac{[H^+][A^-]}{[HA]} \qquad (2)$$

The hydrogen ion concentration $[H^+]$ of a solution of a weak acid can then be calculated as follows. Eq. (2) can be rearranged to give

$$H^+ = K_a \times \frac{[HA]}{[A^-]} \qquad (3)$$

Eq. (3) can be expressed in terms of a negative logarithm:

$$-\log[H^+] = -\log K_a - \log\frac{[HA]}{[A^-]} \qquad (4)$$

Because pH is the negative logarithm of $[H^+]$ (i.e., $-\log[H^+]$), and pK_a equals the negative logarithm of the dissociation constant for a weak acid (i.e., $-\log K_a$), the

Henderson–Hasselbalch Eq. (5) can be developed and used for analysis of acid–base equilibrium systems:

$$\mathrm{pH} = \mathrm{pK_a} + \log\frac{[\mathrm{A^-}]}{[\mathrm{HA}]} \tag{5}$$

For a weak base, such as an amine, the dissociation reaction can be written as

$$\mathrm{RNH_3^+} \rightleftarrows \mathrm{H^+} + \mathrm{RNH_2} \tag{6}$$

and the Henderson–Hasselbalch equation becomes

$$\mathrm{pH} = \mathrm{pK_a} + \log\frac{[\mathrm{RNH_2}]}{[\mathrm{RNH_3^+}]} \tag{7}$$

From Eqs. (5) and (7), it is apparent that the extent of protonation of acidic and basic functional groups, and therefore the net charge of an amino acid, will vary with the pK_a of the functional group and the pH of the solution. For alanine, which has two functional groups with pK_a = 2.4 and 9.8, respectively (Fig. 2.4), the net charge varies with pH, from +1 in a low pH acidic solution to −1 in a high pH basic solution. At a point intermediate between pK_{a1} and pK_{a2}, alanine has a net zero charge. This pH is called its isoelectric point, pI (Fig. 2.4).

BUFFERS

Amino acids and proteins are excellent buffers under physiologic conditions

Buffers are solutions that minimize a change in $[H^+]$ (i.e., pH) on addition of acid or base. A buffer solution, containing a weak acid or weak base and a counter-ion, has maximal buffering capacity at its pK_a - that is, when the acidic and basic forms are present at equal concentrations. The acidic protonated form reacts with added base, and the basic unprotonated form neutralizes added acid, shown as follows for an amino compound:

$$\mathrm{RNH_3^+} + \mathrm{OH^-} \rightleftarrows \mathrm{RNH_2} + \mathrm{H_2O}$$

$$\mathrm{RNH_2} + \mathrm{H^+} \rightleftarrows \mathrm{RNH_3^+}$$

An alanine solution (Fig. 2.4) has maximal buffering capacity at pH 2.4 and 9.8 - that is, at the pK_a of the carboxyl and amino groups, respectively. When dissolved in water, alanine exists as a dipolar ion, or **zwitterion**, in which the carboxyl group is unprotonated ($-COO^-$) and the amino group is protonated ($-NH_3^+$). The pH of the solution is 6.1, the pI (isoelectric point) halfway between the pK_a of the amino and

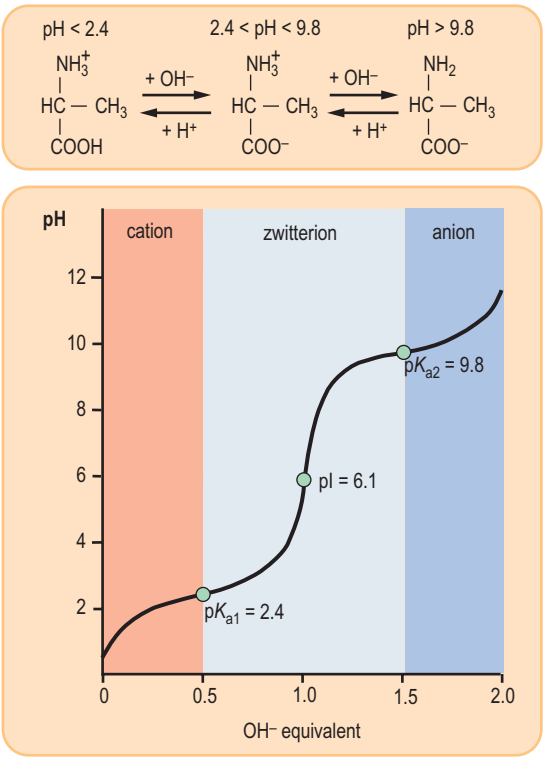
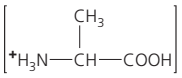
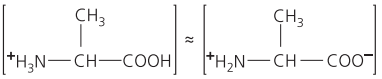


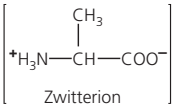
Fig. 2.4 **Titration of amino acid.** The curve shows the number of equivalents of NaOH consumed by alanine while titrating the solution from pH 0 to pH 12. Alanine contains two ionizable groups: an α-carboxyl group and an α-amino group. As NaOH is added, these two groups are titrated. The pK_a of the α-COOH group is 2.4, whereas that of the α-NH_3^+ group is 9.8. At very low pH, the predominant ion species of alanine is the fully protonated, cationic form:



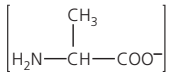
At the midpoint in the first stage of the titration (pH 2.4), equimolar concentrations of proton donor and proton acceptor species are present, providing good buffering power.



At the midpoint in the overall titration (pH 6.1), the zwitterion is the predominant form of the amino acid in solution. The amino acid has a net zero charge at this pH, the negative charge of the carboxylate ion being neutralized by the positive charge of the ammonium group.



The second stage of the titration corresponds to the removal of a proton from the $-NH_3^+$ group of alanine. The pH at the midpoint of this stage is 9.8, equal to the pK_a for the $-NH_3^+$ group. The titration is complete at a pH of about 12, at which point the predominant form of alanine is the unprotonated, anionic form:



The pH at which a molecule has no net charge is known as its isoelectric point, pI. For alanine, it is calculated as follows:

$$\mathrm{pI} = \frac{\mathrm{pK_{a1}} + \mathrm{pK_{a2}}}{2} = \frac{(2.4 + 9.8)}{2} = 6.1$$

carboxyl groups. The titration curve of alanine by NaOH (Fig. 2.4) illustrates that alanine has minimal buffering capacity at its pI and maximal buffering capacity at a pH equal to its pK_{a1} or pK_{a2}.

PEPTIDES AND PROTEINS

Primary structure of proteins

The primary structure of a protein is the linear sequence of its amino acids

In proteins, the carboxyl group of one amino acid is linked to the amino group of the next amino acid, forming an amide (peptide) bond; water is eliminated during the reaction (Fig. 2.5). The amino acid units in a peptide chain are referred to as amino acid residues. Peptide chains consisting of three or more amino acid residues are called a tripeptide (four are a tetrapeptide, five are a pentapeptide, and so on; all are polypeptides). By convention, the amino terminus (N-terminus) is taken as the first residue, and the sequence of amino acids is written from left to right. When writing the peptide sequence, one uses either the three-letter or the one-letter abbreviations for amino acids, such as Asp-Arg-Val-Tyr-Ile-His-Pro-Phe-His-Leu or D-R-V-Y-I-H-P-F-H-L (Table 2.1); this peptide is angiotensin, a decapeptide hormone that affects blood pressure. The amino acid residue having a free amino group at one end of the peptide, Asp, is called the N-terminal amino acid (amino terminus), whereas the residue having a free carboxyl group at the other end, Leu, is called the C-terminal amino acid (carboxyl terminus). Proteins contain between 50 and 2000 amino acid residues. The mean molecular mass of an amino acid residue is about 110 dalton units (Da), and the molecular mass of most proteins is between 5500 and 220,000 Da.

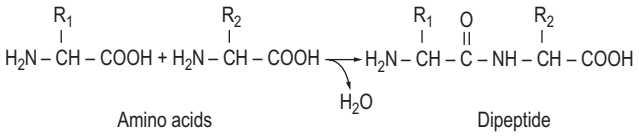


Fig. 2.5 **Structure of a peptide bond.**

Amino acid side chains contribute both charge and hydrophobicity to proteins

The amino acid composition of a peptide chain has a profound effect on its physical and chemical properties. Proteins rich in aliphatic or aromatic amino groups are relatively insoluble in water and are likely to be found in cell membranes. Proteins rich in polar amino acids are more water soluble; they are found in the cellular cytoplasm and vacuolar compartments, and in blood plasma and extravascular fluids. Amides are neutral compounds, so the amide backbone of a protein, including the α-amino and α-carboxyl groups from which it is formed, does not contribute to the charge of the protein. Instead, the charge on the protein depends primarily on the side chain amino and carboxyl functional groups of amino acids, plus a minor contribution from the amino and carboxyl groups of the terminal amino acids. Amino acids with side chain acidic (Glu, Asp) or basic (Lys, His, Arg) groups confer both charge and buffering capacity to a protein. The balance between acidic and basic side chains in a protein determines the protein's **isoelectric point** and its net charge in solution. Proteins rich in lysine and arginine are basic in solution and have a positive charge at neutral pH, whereas proteins rich in acidic aspartate and glutamate are acidic and have a negative charge at neutral pH. Because of their side chain functional groups, all proteins become more positively charged at acidic pH and more negatively charged at basic pH. Proteins are an important part of the buffering capacity of cells and biological fluids, including blood; they have maximum buffering capacity at their isoelectric point (pI).

Secondary structure of proteins

The secondary structure of a protein is determined by hydrogen bond interactions between backbone carbonyl and amide groups

The secondary structure of a protein refers to the local structure of the polypeptide chain. This structure is determined by hydrogen bond interactions between the carbonyl oxygen group of one peptide bond and the amide hydrogen of another nearby peptide bond. There are two types of secondary structure: the α-helix and the β-pleated sheet.

The α-helix

The α-helix is a rodlike structure, with the peptide chain tightly coiled and the side chains of amino acid residues extending outward from the axis of the spiral. Each amide carbonyl group is hydrogen-bonded to the amide hydrogen of a peptide bond that is four residues away along the same chain. There are, on average, 3.6 amino acid residues per turn of the helix, and the helix winds in a right-handed (clockwise) manner in almost all proteins (Fig. 2.6A).

The β-pleated sheet

If the H-bonds are formed laterally between peptide bonds, the polypeptide sequences become arrayed parallel or antiparallel to one another in what is commonly called a β-pleated sheet. The β-pleated sheet is an extended structure as opposed to the coiled α-helix. It is pleated because the carbon–carbon (C–C) bonds are tetrahedral and cannot exist in a planar configuration. If the polypeptide chains run in the same direction, they form a parallel β-sheet (Fig. 2.6B), but in the opposite

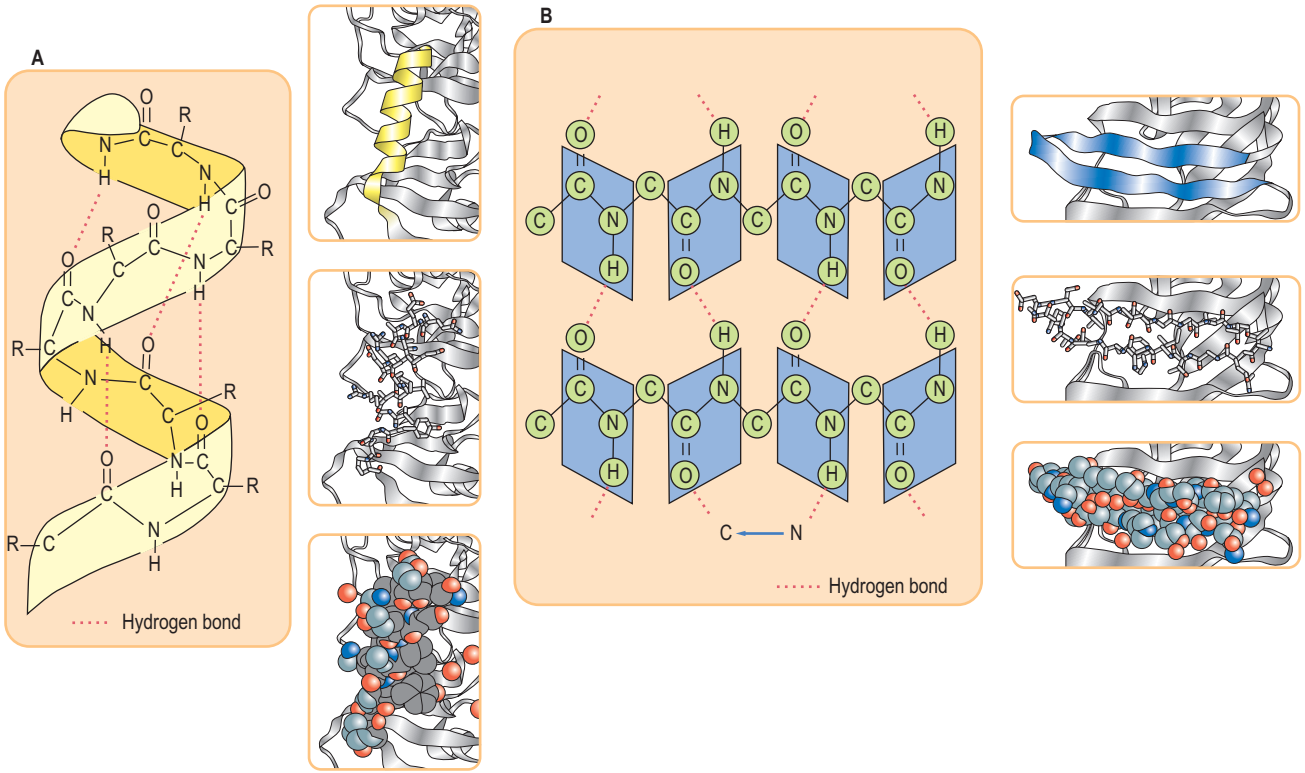


Fig. 2.6 **Protein secondary structural motifs.** (A) An α-helical secondary structure. Hydrogen bonds between "backbone" amide NH and C=O groups stabilize the α-helix. Hydrogen atoms of the OH, NH, or SH group (hydrogen donors) interact with electron pairs of the acceptor atoms such as O, N, or S. Even though the bonding energy is lower than that of covalent bonds, hydrogen bonds play a pivotal role in the stabilization of protein molecules. Ribbon, stick, and space-filling models are shown. R represents a side chain of amino acids that extends outward from the helix. (B) The parallel β-sheet secondary structure. In the β-conformation, the backbone of the polypeptide chain is extended into a zigzag structure. When the zigzag polypeptide chains are arranged side by side, they form a structure resembling a series of pleats. Ribbon, stick, and space-filling models are shown.

direction, they form an antiparallel structure. The β-turn, or β-bend, refers to the segment in which the polypeptide abruptly reverses direction. Glycine (Gly) and proline (Pro) residues often occur in β-turns on the surface of globular proteins.

Tertiary structure of proteins

The tertiary structure of a protein is determined by interactions between side chain functional groups, including disulfide bonds, hydrogen bonds, salt bridges, and hydrophobic interactions

The three-dimensional, folded, and biologically active conformation of a protein is referred to as its tertiary structure. This structure reflects the overall shape of the molecule and generally consists of several smaller folded units termed **domains**.

The tertiary structure of a protein is stabilized by interactions between side chain functional groups: covalent disulfide bonds, hydrogen bonds, salt bridges, and hydrophobic interactions (Fig. 2.7). The side chains of tryptophan and arginine serve as hydrogen donors, whereas asparagine, glutamine,

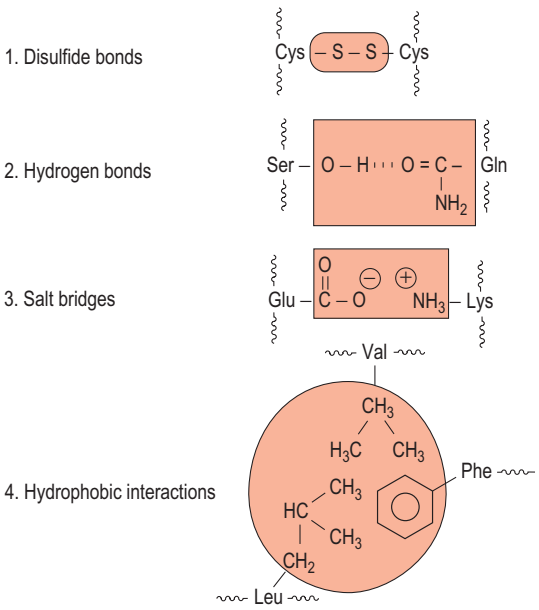


Fig. 2.7 **Elements of tertiary structure of proteins.** Examples of amino acid side chain interactions contributing to tertiary structure.

serine, and threonine can serve as both hydrogen donors and acceptors. Lysine, aspartic acid, glutamic acid, tyrosine, and histidine also can serve as both donors and acceptors in the formation of ion pairs (salt bridges). Two oppositely charged amino acids, such as glutamate with a γ-carboxyl group and lysine with an ε-amino group, may form a salt bridge, primarily on the surface of proteins. Compounds such as urea and guanidine hydrochloride block these interactions and cause denaturation, or loss of secondary and tertiary structure, when present at high concentrations for example, 8 mol/L urea. These reagents are called **denaturants** or **chaotropic agents.**

ADVANCED CONCEPT BOX
COLLAGEN

Human genetic defects involving collagen illustrate the close relationship between amino acid sequence and three-dimensional structure. Collagens are the most abundant protein family in the mammalian body, representing about a third of body protein. Collagens are a major component of connective tissue such as cartilage, tendons, the organic matrix of bones, and the cornea of the eye.

Comment

Collagen contains 35% Gly, 11% Ala, and 21% Pro plus Hyp (hydroxyproline). The amino acid sequence in collagen is generally a repeating tripeptide unit, Gly-Xaa-Pro or Gly-Xaa-Hyp, where Xaa can be any amino acid; Hyp = hydroxyproline. In contrast to the α-helical structure of other proteins, collagen forms a left-handed helical structure with three residues per turn. Three of these helices wrap around one another with a right-handed twist. The resulting triple-stranded molecule is referred to as tropocollagen. Tropocollagen molecules self-assemble into collagen fibrils and are packed together to form collagen fibers. There are metabolic and genetic disorders that result from collagen abnormalities. Scurvy, osteogenesis imperfecta (Chapter 19), and Ehlers–Danlos syndrome result from defects in collagen synthesis and/or crosslinking.

Quaternary structure of proteins

The quaternary structure of multisubunit proteins is determined by covalent and noncovalent interactions between the subunit surfaces

Quaternary structure refers to a complex, or an assembly, of two or more separate peptide chains that are held together by noncovalent or, in some cases, covalent interactions. In general, most proteins larger than 50 kDa consist of more than one chain and are referred to as dimeric, trimeric, or multimeric proteins. Many multisubunit proteins are composed of different kinds of **functional subunits, such as the regulatory and catalytic subunits**. Hemoglobin is a tetrameric protein (Chapter 5), and beef heart mitochondrial ATPase has 10 protomers (Chapter 8). The smallest unit is referred to as a monomer, or subunit. Fig. 2.8 illustrates the structure of the dimeric protein

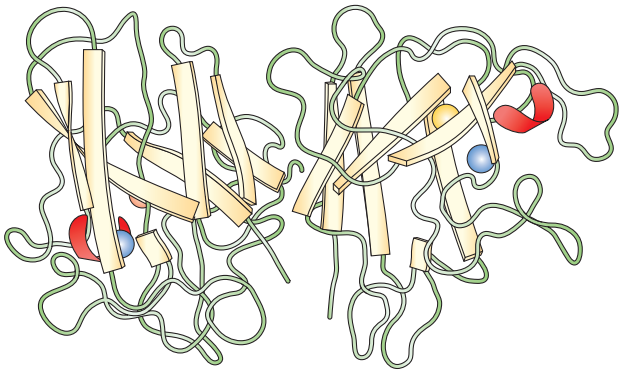


Fig. 2.8 **Three-dimensional structure of a dimeric protein.** Quaternary structure of Cu,Zn-superoxide dismutase from spinach. Cu,Zn-superoxide dismutase has a dimeric structure, with a monomer molecular mass of 16,000 Da. Each subunit consists of eight antiparallel β-sheets called a β-barrel structure, in analogy with geometric motifs found on Native American and Greek weaving and pottery. Red arc, short a-helix. Courtesy Dr. Y. Kitagawa.

CLINICAL BOX
CREUTZFELDT-JAKOB DISEASE

A 56-year-old male cattle rancher presented with epileptic cramps and dementia and was diagnosed as having Creutzfeldt–Jakob disease, a human prion disease. The **prion diseases**, also known as transmissible spongiform encephalopathies, are neurodegenerative diseases that affect both humans and animals. This disease in sheep and goats is designated as scrapie and in cows as spongiform encephalopathy (mad cow disease). The diseases are characterized by the accumulation of an abnormal isoform of a host-encoded protein, prion protein-cellular form (PrPC), in affected brains.

Comment

Prions appear to be composed only of PrPSc (scrapie form) molecules, which are abnormal conformers of the normal, host-encoded protein. PrPC has a high α-helical content and is devoid of β-pleated sheets, whereas PrPSc has a high β-pleated sheet content. The conversion of PrPC into PrPSc involves a profound conformational change. The progression of infectious prion diseases appears to involve an interaction between PrPC and PrPSc that induces a conformational change of the α-helix-rich PrPC to the β-pleated sheet–rich conformer of PrPSc. PrPSc-derived prion disease may be genetic or infectious. The amino acid sequences of different mammalian PrPCs are similar, and the conformation of the protein is virtually the same in all mammalian species.

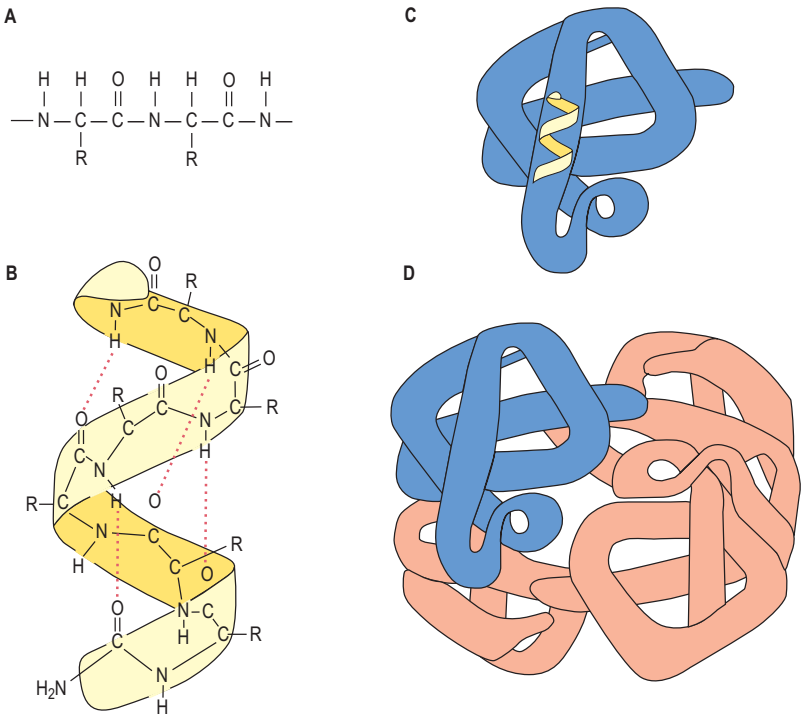


Fig. 2.9 **Primary, secondary, tertiary, and quaternary structures.** (A) The primary structure is composed of a linear sequence of amino acid residues of proteins. (B) The secondary structure indicates the local spatial arrangement of the polypeptide backbone yielding an extended α-helical or β-pleated sheet structure, as depicted by the ribbon. Hydrogen bonds between the "backbone" amide NH and C=O groups stabilize the helix. (C) The tertiary structure illustrates the three-dimensional conformation of a subunit of the protein, and the quaternary structure (D) indicates the assembly of multiple polypeptide chains into an intact, tetrameric protein.

Cu,Zn-superoxide dismutase. Fig. 2.9 is an overview of the primary, secondary, tertiary, and quaternary structures of a tetrameric protein.

ADVANCED CONCEPT BOX
PROTEIN FOLDING

For proteins to function properly, they must fold into the correct shape, or conformation. Proteins have evolved so that one conformation is more favorable than all others: the native state. Numerous proteins assist other proteins in the folding process. These proteins, termed **chaperones**, include "heat shock" proteins, such as HSP 60 and HSP 70, and protein disulfide isomerases. A protein-folding disease is a disease that is associated with the abnormal conformation of a protein. This occurs in chronic, age-related diseases, such as Alzheimer's disease, amyotrophic lateral sclerosis (ALS: Lou Gehrig's disease), and Parkinson's disease. The accumulation of aggregates of misfolded protein contributes to the development of pathology in these diseases.

PURIFICATION AND CHARACTERIZATION OF PROTEINS

Protein purification is a multistep process, based on protein size, charge, solubility, and ligand binding

The complete characterization of a protein requires its purification and determination of its complete primary, secondary, and tertiary structure and, for a multimeric protein, its quaternary structure. To characterize a protein, it is first necessary to purify the protein by separating it from other components in complex biological mixtures. The source of the proteins is commonly blood or tissues or microbial cells such as bacteria and yeast. First, the cells or tissues are disrupted by grinding or homogenization in buffered isotonic solutions, commonly at physiologic pH and at 4°C to minimize protein denaturation during purification. The "crude extract" containing organelles such as nuclei, mitochondria, lysosomes, microsomes, and cytosolic fractions can then be fractionated by high-speed centrifugation, or ultracentrifugation. Proteins that are tightly bound to other biomolecules or membranes may be solubilized using organic solvent or detergent.

Protein purification-precipitation

Protein purification is based on differences in a protein's solubility, size, charge, and binding properties

The solubility of a protein may be increased by the addition of salt at a low concentration (salting in) or decreased by high salt concentration (salting out). When ammonium sulfate, one of the most soluble salts, is added to a solution of a protein, some proteins precipitate at a given salt concentration, whereas others do not. Human serum immunoglobulins are precipitable by 33%–40% saturated $(NH_4)_2SO_4$, whereas serum albumin remains soluble. Saturated ammonium sulfate is about 4.1 mol/L. Most proteins will precipitate from an 80% saturated $(NH_4)_2SO_4$ solution.

Proteins may also be precipitated from solution by adjusting the pH. Proteins are generally least soluble at their isoelectric point (pI). At this pH, the protein has no net charge or charge–charge repulsion between subunits. Hydrophobic interactions between protein surfaces then lead to aggregation and precipitation of the protein.

Dialysis and ultrafiltration

Small molecules, such as salts, can be removed from protein solutions by dialysis or ultrafiltration

Dialysis is performed by adding the protein–salt solution to a semipermeable membrane tube (commonly a nitrocellulose or collodion membrane). When the tube is immersed in a dilute buffer solution, small molecules will pass through, but large protein molecules will be retained in the tube, depending on the pore size of the dialysis membrane. This procedure is particularly useful for removal of $(NH_4)_2SO_4$ or other salts during protein purification because the salts will interfere with the purification of proteins by ion-exchange chromatography (below). Fig. 2.10 illustrates the dialysis of proteins.

Ultrafiltration has largely replaced dialysis for purification of proteins. This technique uses pressure to force a solution through a semipermeable membrane of a defined, homogeneous pore size. By selecting the proper molecular-weight cut-off (MWCO; pore size) value for the filter, the membranes will allow solvent and lower-molecular-weight solutes to pass through the membrane, forming the filtrate, while retaining higher-molecular-weight proteins in the retentate solution. Ultrafiltration can be used to concentrate protein solutions or to accomplish dialysis by continuous replacement of buffer in the retentate compartment.

Gel filtration (molecular sieving)

Gel filtration chromatography separates proteins on the basis of size

Gel filtration, or gel-permeation or size-exclusion, chromatography uses a column of insoluble but highly hydrated

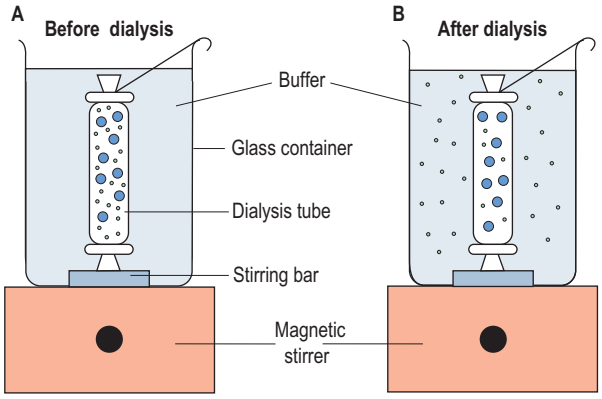


Fig. 2.10 **Dialysis of proteins.** Protein and low-molecular-mass compounds are separated by dialysis on the basis of size. (A) A protein solution with salts is placed in a dialysis tube in a beaker and dialyzed with stirring against an appropriate buffer. (B) The protein is retained in the dialysis tube, whereas salts will exchange through the membrane. By use of a large volume of external buffer, with occasional buffer replacement, the protein will eventually be exchanged into the external buffer solution.

polymers, such as dextran, agarose, or polyacrylamide. Gel filtration chromatography depends on the differential migration of dissolved solutes through gels that have pores of defined sizes. This technique is frequently used for protein purification and for desalting protein solutions. Fig. 2.11 describes the principle of gel filtration. There are commercially available gels made from polymer beads designated as dextran (Sephadex series), polyacrylamide (Bio-Gel P series), and agarose (Sepharose series), respectively. The gels vary in pore size, and one can choose the gel filtration materials according to the molecular-weight-fractionation range desired.

Ion-exchange chromatography

Proteins bind to ion-exchange matrices based on charge–charge interactions

When a charged ion or molecule with one or more positive charges exchanges with another positively charged component bound to a negatively charged immobilized phase, the process is called cation exchange. The inverse process is called anion exchange. The cation exchanger, carboxymethylcellulose ($R–O–CH_2–COO^-$), and anion exchanger, diethylaminoethyl (DEAE) cellulose [$R–O–C_2H_4–NH^+(C_2H_5)_2$], are frequently used for the purification of proteins. Consider purifying a protein mixture containing albumin and immunoglobulin. At pH 7.5, albumin, with a pI of 4.8, is negatively charged; immunoglobulin, with a pI ~8.0, is positively charged. If the mixture is applied to a DEAE-cellulose column at pH 7.0, the albumin sticks to the positively charged column, whereas the immunoglobulin passes through the column. Fig. 2.12 illustrates the principle

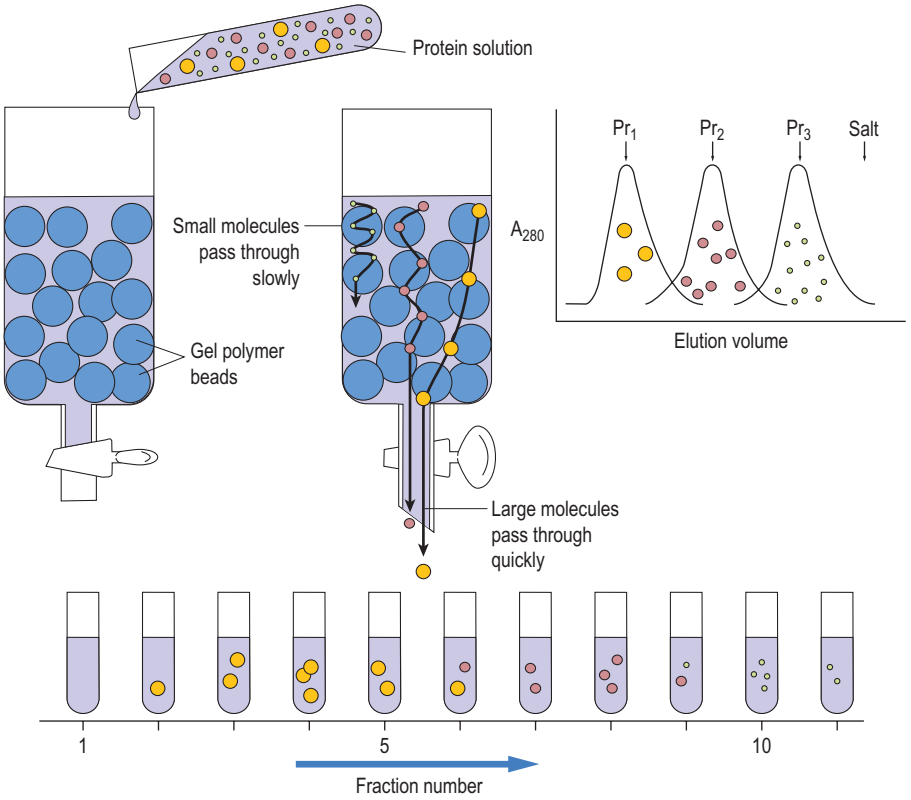


Fig. 2.11 **Fractionation of proteins by size: gel filtration chromatography of proteins.** Proteins with different molecular sizes are separated by gel filtration based on their relative size. The smaller the protein, the more readily it exchanges into polymer beads, whereas larger proteins may be completely excluded. Larger molecules flow more rapidly through this column, leading to fractionation on the basis of molecular size. The chromatogram on the right shows a theoretical fractionation of three proteins, Pr_1–Pr_3, of decreasing molecular weight.

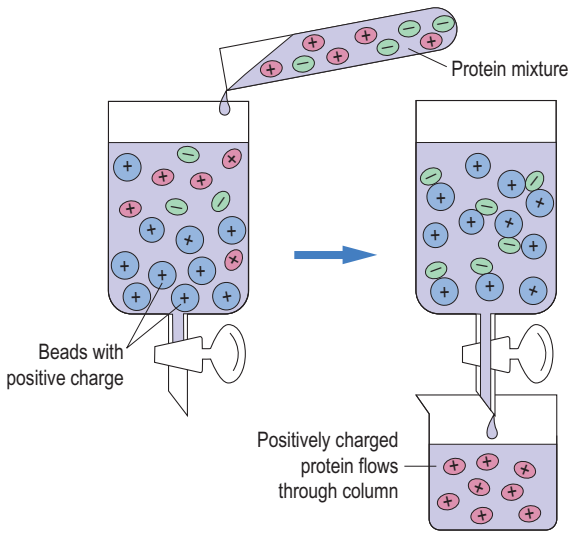


Fig. 2.12 **Fractionation of proteins by charge: ion-exchange chromatography.** Mixtures of proteins can be separated by ion-exchange chromatography according to their net charges. Beads that have positively charged groups attached are called anion exchangers, whereas those having negatively charged groups are cation exchangers. This figure depicts an anion-exchange column. Negatively charged protein binds to positively charged beads, and positively charged protein flows through the column.

ADVANCED CONCEPT BOX
HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY (HPLC)

HPLC is a powerful chromatographic technique for high-resolution separation of proteins, peptides, and amino acids. The principle of the separation may be based on the charge, size, or hydrophobicity of proteins. The narrow columns are packed with a noncompressible matrix of fine silica beads coated with a thin layer of a stationary phase. A protein mixture is applied to the column, and then the components are eluted by either isocratic or gradient chromatography. The eluates are monitored by ultraviolet absorption, refractive index, or fluorescence. This technique uses finely packed micron-size beads and requires high pressure for efficient elution, but it yields high-resolution separations.

of ion-exchange chromatography. As with gel-permeation chromatography, proteins can be separated from one another based on small differences in their pI. **Adsorbed proteins are commonly eluted with a gradient formed from two or more solutions with different pH and/or salt concentrations**. In this way, proteins are gradually eluted from the column and are separated based on their pI.

Affinity chromatography

Affinity chromatography purifies proteins based on ligand interactions

Affinity chromatography is a convenient and specific method for the purification of proteins. A porous chromatography column matrix is derivatized with a ligand that interacts with, or binds to, a specific protein in a complex mixture. The protein of interest will be selectively and specifically bound to the ligand, whereas the others wash through the column. The bound protein can then be eluted by a high salt concentration, by mild denaturation, or by a soluble form of the ligand or ligand analogs.

Determination of purity and molecular weight of proteins

Polyacrylamide gel electrophoresis in sodium dodecylsulfate can be used to separate proteins based on size

Electrophoresis can be used for the separation of a wide variety of charged molecules, including amino acids, polypeptides, proteins, and DNA. When a current is applied to molecules in dilute buffers, those with a net negative charge at the selected pH migrate toward the anode, and those with a net positive charge migrate toward the cathode. A porous support, such as paper, cellulose acetate, or a polymeric gel, is commonly used to minimize diffusion and convection.

Like chromatography, electrophoresis may be used for preparative fractionation of proteins at physiologic pH. Different soluble proteins will move at different rates in the electrical field, depending on their charge-to-mass ratio. A denaturing detergent, sodium dodecylsulfate (SDS), is commonly used in a polyacrylamide gel electrophoresis (PAGE) system to separate and resolve protein subunits according to molecular weight. The protein preparation is usually treated with both SDS and a thiol reagent, such as β-mercaptoethanol, to reduce disulfide bonds and unfold the peptide chain into a random coil. Because the binding of SDS is proportional to the length of the peptide chain, each protein molecule has the same mass-to-charge ratio, and the relative mobility of the protein in the polyacrylamide matrix is proportional to the molecular mass of the polypeptide chain. Varying the extent of crosslinking of the polyacrylamide gel provides selectivity for separation of proteins with different ranges of molecular weights. A purified protein preparation can be readily analyzed for homogeneity on SDS-PAGE by staining with dyes, such as Coomassie Blue, or with a silver staining technique, as shown in Fig. 2.13.

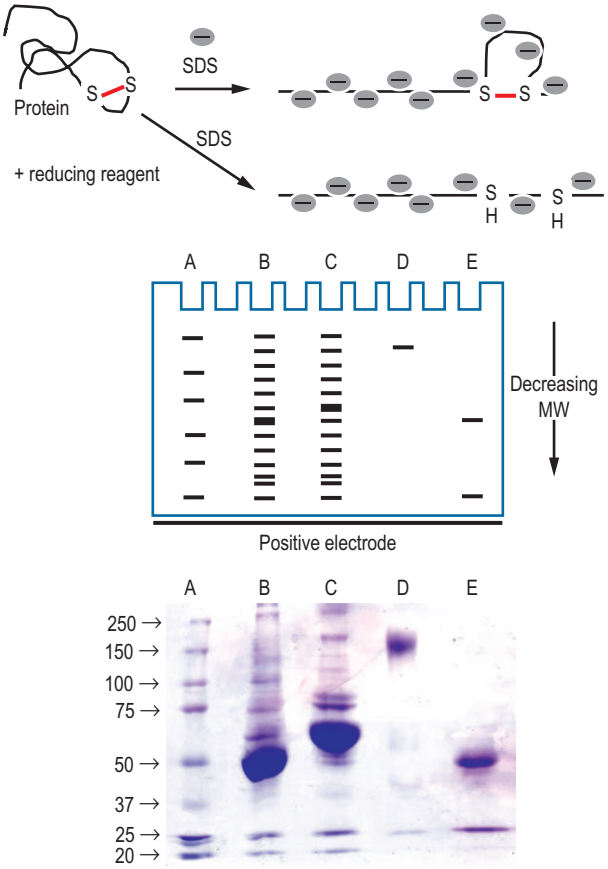


Fig. 2.13 **SDS-PAGE.** Sodium dodecylsulfate (SDS) polyacrylamide gel electrophoresis (PAGE) is used to separate proteins on the basis of their molecular weights. SDS is an anionic detergent which binds to and denatures proteins and imparts a negative charge to protein molecules. However, intra-molecular crosslinking by disulfide bond between 2 cysteines is still intact. SDS treatment of proteins in the presence of reducing reagents such as 2-mercaptoethanol cleaves disulfide bonds, generating a linearized protein with similar charge-to-mass ratios. Larger molecules are retarded in the gel matrix, whereas the smaller ones move more rapidly. Lane A contains standard proteins with known molecular masses (indicated in kDa on the left). Lanes B, C, D, and E show results of SDS-PAGE analysis of a protein at various stages in purification: B, total serum proteins; C, total serum proteins with reducing reagent; D, purified immunoglobulin; E, purified immunoglobulin with reducing reagent.

Isoelectric focusing (IEF)

IEF resolves proteins based on their isoelectric point

IEF is conducted in a microchannel or gel containing a stabilized pH gradient, formed by ampholytes, which are zwitterionic species with a range of isoelectric points. When a charge is applied to the solution, the ampholytes self-organize into a stable pH gradient. A protein applied to the system will be either positively or negatively charged, depending on its amino acid composition and the ambient pH. Upon application of a current, the protein will move toward either the anode or cathode until it encounters that part of the system that corresponds to its pI, where the protein has no charge and will cease to migrate. **IEF is used in conjunction with SDS-PAGE for two-dimensional gel electrophoresis** (Fig. 2.14). This technique is particularly useful for the fractionation of complex mixtures of proteins for proteomic analysis.

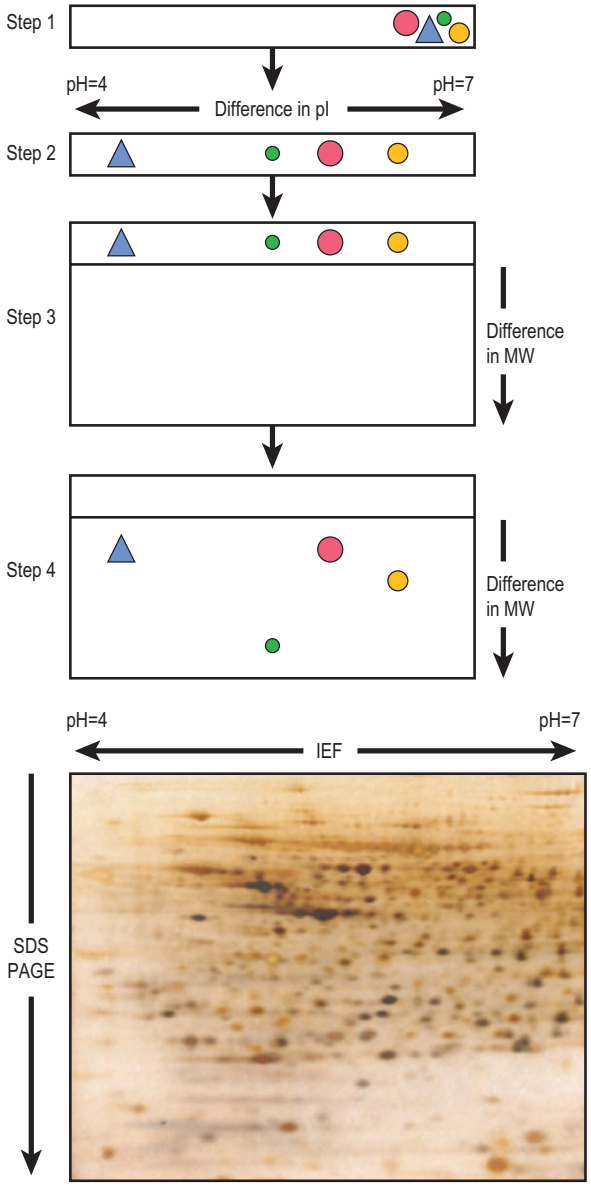


Fig. 2.14 **Two-dimensional gel electrophoresis.** (Top) **Step 1:** Sample containing proteins is applied to a cylindrical isoelectric-focusing gel within the pH gradient. **Step 2:** Each protein migrates to a position in the gel corresponding to its isoelectric point (pI). **Step 3:** The IEF gel is placed horizontally on the top of a slab gel. **Step 4:** The proteins are separated by SDS-PAGE according to their molecular weight. (Bottom) Typical example of 2D-PAGE. A rat liver homogenate was fractionated by 2D-PAGE, and proteins were detected by silver staining.

ANALYSIS OF PROTEIN STRUCTURE

The typical steps in the purification of a protein are summarized in Fig. 2.15. Once purified, for the determination of its amino acid composition, a protein is subjected to hydrolysis, commonly in 6 mol/L HCl at 110°C in a sealed and evacuated

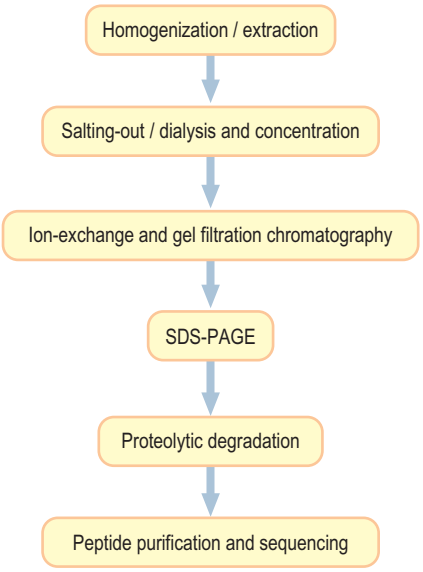


Fig. 2.15 **Strategy for protein purification.** Purification of a protein involves a sequence of steps in which contaminating proteins are removed based on differences in size, charge, and hydrophobicity. Purification is monitored by SDS-PAGE (Fig. 2.13).

tube for 24–48 h. Under these conditions, tryptophan, cysteine, and most of the cystine are destroyed, and glutamine and asparagine are quantitatively deaminated to give glutamate and aspartate, respectively. Recovery of serine and threonine is incomplete and decreases with increasing time of hydrolysis. Alternative hydrolysis procedures may be used for measurement of tryptophan, whereas cysteine and cystine may be converted to an acid-stable cysteic acid prior to hydrolysis.

After hydrolysis, the free amino acids are separated on an automated amino acid analyzer using an ion-exchange column or, after pre-column derivatization with colored or fluorescent reagents, by reversed-phase high-performance (hydrophobic surface) liquid chromatography (RP-HPLC). Free amino acids fractionated by ion-exchange chromatography are detected by post-column reaction with a chromogenic or fluorogenic reagent, such as ninhydrin or dansyl chloride, Edman's reagent (below), or *o*-phthalaldehyde. These techniques allow the measurement of as little as 1 pmol of each amino acid. A typical elution pattern of amino acids in a purified protein is shown in Fig. 2.16.

Determination of the primary structure of proteins

Historically, analysis of protein sequence was carried out by chemical methods; today, both sequence analysis and protein identification are performed by mass spectrometry

Information on the primary sequence of a protein is essential for understanding its functional properties, the identification of the family to which the protein belongs, and the

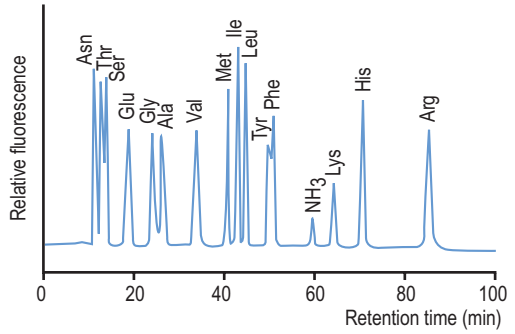


Fig. 2.16 **Chromatogram from an amino acid analysis by cation-exchange chromatography.** A protein hydrolysate is applied to the cation-exchange column in a dilute buffer at acidic pH (~3.0), at which all amino acids are positively charged. The amino acids are then eluted by a gradient of increasing pH and salt concentrations. The most anionic (acidic) amino acids elute first, followed by the neutral and basic amino acids. Amino acids are visualized by post-column reaction with a fluorogenic compound, such as *o*-phthalaldehyde.

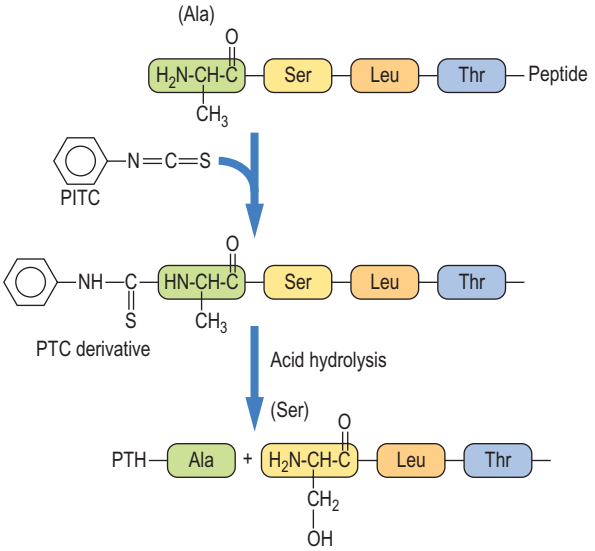


Fig. 2.17 **Steps in Edman degradation.** The Edman degradation method sequentially removes one residue at a time from the amino end of a peptide. Phenyl isothiocyanate (PITC) converts the N-terminal amino group of the immobilized peptide to a phenylthiocarbamyl (PTC) amino acid derivative in alkaline solution. Mild acid treatment removes the first amino acid as the phenylthiohydantoin (PTH) derivative, which is identified by HPLC.

ADVANCED CONCEPT BOX
THE PROTEOME

A proteome is defined as the full complement of proteins produced by a particular genome. Proteomics is defined as the qualitative and quantitative comparison of proteomes under different conditions. The proteome is tissue and cell specific, and it changes during development and in response to hormonal signaling and environmental stresses. In one approach to analyzing the proteome of a cell, proteins are extracted and subjected to two-dimensional (2D) polyacrylamide gel electrophoresis (2D-PAGE; Fig. 2.14). Individual protein spots are identified by staining, then extracted and digested with proteases. Small peptides from such a gel are sequenced by mass spectrometry, permitting the identification of the protein. In 2D-differential gel electrophoresis (DIGE), two proteomes may be compared by labeling their proteins with different fluorescent dyes (e.g., red and green). The labeled proteins are mixed, then fractionated by 2D-PAGE. Proteins present in both proteomes will appear as yellow spots, whereas unique proteins will be red or green, respectively (Chapter 24).

characterization of mutant proteins that cause disease. Because of the large size of proteins, they are typically cleaved by digestion by specific endoproteases, such as trypsin (Chapter 6), V8 protease, or lysyl endopeptidase, to obtain peptide fragments. Trypsin cleaves peptide bonds on the C-terminal side of arginine and lysine residues, provided the next residue is not proline. Lysyl endopeptidase is also frequently used to cleave at the C-terminal side of lysine. Cleavage by chemical reagents such as cyanogen bromide is also useful; cyanogen bromide cleaves on the C-terminal side of methionine residues. Before cleavage, proteins with cysteine and cystine residues are reduced by 2-mercaptoethanol and then treated with iodoacetate to form carboxymethylcysteine residues. This avoids spontaneous formation of inter- or intramolecular disulfides during analyses.

The cleaved peptides are then subjected to RP-HPLC to purify the peptide fragments, then sequenced on an automated protein sequencer using the **Edman degradation** technique (Fig. 2.17). The sequence of overlapping peptides prepared by different cleavage reagents is then used to obtain the primary structure of the protein. The Edman degradation technique is largely of historical interest. Mass spectrometry is more commonly used today to obtain both the molecular mass and sequence of polypeptides simultaneously (Chapter 24). Both techniques can be applied directly to proteins or peptides recovered from SDS-PAGE or 2D-PAGE.

Protein sequencing and identification is currently done by electrospray ionization liquid chromatography tandem mass spectrometry (HPLC-ESI-MS/MS; Chapter 24). This technique is sufficiently sensitive that proteins separated by 2D-PAGE (Fig. 2.14) can be recovered from the gel for analysis. As little as 1 µg of protein can be digested with trypsin in situ, then extracted from the gel and identified based on the amino acid sequence of its peptides. This technique, and a complementary technique called matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) MS/MS (Chapter 24), can be applied for determination of the molecular weight of intact proteins and for sequence analysis of peptides, leading to unambiguous identification of a protein.

Determination of the three-dimensional structure of proteins

X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy are usually used for determination of the three-dimensional structure of proteins

X-ray crystallography depends on the diffraction of x-rays by the electrons of the atoms constituting the molecule. However, because the x-ray diffraction caused by an individual molecule is weak, the protein must exist in the form of a well-ordered crystal, in which each molecule has the same conformation in a specific position and orientation on a three-dimensional lattice. Based on diffraction of a collimated beam of electrons, the distribution of the electron density, and thus the location of atoms, in the crystal can be calculated to determine the structure of the protein. For protein crystallization, the most frequently used method is the hanging drop method, which involves the use of a simple apparatus that permits a small portion of a protein solution (typically a 10-μL droplet containing 0.5–1.0 mg/protein) to evaporate gradually to reach the saturation point at which the protein begins to crystallize. NMR spectroscopy is commonly used for structural analysis of small organic compounds, but high-field NMR is also useful for determination of the structure of a protein in solution and complements information obtained by x-ray crystallography.

SUMMARY

- Proteins are macromolecules formed by polymerization of amino acids. There are 20 different α-L-amino acids in proteins, linked by peptide bonds. The side chains of the amino acids contribute charge, polarity, and hydrophobicity to proteins.
- The linear sequence of the amino acids constitutes the primary structure of the protein. Higher-order structures are formed by hydrogen bonds between backbone carbonyl and amide groups (secondary structure), by hydrophobic interactions, by salt bridges and covalent bonds between the side chains of amino acids (tertiary structure), and by noncovalent association of multiple polypeptide chains to form polymeric proteins (quaternary structure).
- Purification and characterization of proteins are essential for elucidating their structure and function. By taking advantage of differences in their solubility, size, charge, and ligand-binding properties, proteins can be purified to homogeneity using various chromatographic and electrophoretic techniques. The molecular mass and purity of a protein, and its subunit composition, can be determined by SDS-PAGE.
- Deciphering the primary and three-dimensional structures of a protein by chemical methods, mass spectrometry, x-ray crystallographic analysis, and NMR spectroscopy leads to an understanding of structure–function relationships in proteins.

ACTIVE LEARNING

1. Mass spectrometry analysis of blood, urine, and tissues is now being applied for clinical diagnosis. Discuss the merits of this technique with respect to specificity, sensitivity, throughput, and breadth of analysis, including proteomic analysis for diagnostic purposes.
2. Review the importance of protein misfolding and deposition in tissues in age-related chronic diseases.

FURTHER READING

Bada JL. New insights into prebiotic chemistry from Stanley Miller's spark discharge experiments. *Chemical Society Reviews*. 2013;42:2186–2196.

Dhingra S, Sowdhamini R, Cadet F, Offman B. A glance into the evolution of template-free protein structure prediction methodologies. *Biochimie*. 2020;175:85–92.

Dill KA, MacCallum JL. The protein-folding problem, 50 years on. *Science*. 2012;338:1042–1046. http://science.sciencemag.org/content/338/6110/1042.full.

Li ZL, Buck M. Beyond history and "on a roll": The list of the most well-studied human protein structures and overall trends in the Protein Data Bank. *Protein Science*. 2021:7. Feb. https://doi.org/10.1002/pro.4038.

Raoufinia R, Mota A, Keyhanvar N, et al. Overview of albumin and its purification methods. *Advanced Pharmaceutical Bulletin*. 2016;6:495–507.

Simien JM, Haglund E. Topological twists in Nature. *Trends in Biochemical Science*. 2021, in press. https://doi.org/10.1016/j.tibs.2020.12.004.

Watts JC, Prusiner SB. β-Amyloid prions and the pathobiology of Alzheimer's disease. *Cold Spring Harbor Perspectives in Medicine*. 2018. http://doi.org/10.1101/cshperspect.a023507. PMID: 28193770.

RELEVANT WEBSITES

Protein Data Banks: http://www.rcsb.org
https://www.ncbi.nlm.nih.gov/protein/
YouTube and Khan Academy videos on protein structure and folding.

In addition:
Protein structure: https://www.youtube.com/watch?v=1peFJ_-N7V8&ab_channel=NeuralAcademy
YouTube: Protein folding by AI/DeepMind
TED Talk: The protein folding problem: a major conundrum of science: https://www.youtube.com/watch?v=zm-3kovWpNQ&ab_channel=TEDxTalks

ABBREVIATIONS

GSH	Glutathione (reduced)
GSSG	Glutathione (oxidized)
HPLC	High performance (pressure) liquid chromatography
IEF	Isoelectric focusing
MWCO	Molecular weight cut-off
PAGE	Polyacrylamide gel electrophoresis
pI	Isoelectric point
RP-HPLC	Reversed phase HPLC
SDS	Sodium dodecylsulfate
SDS-PAGE	Sodium dodecylsulfate-polyacrylamide gel electrophoresis

CHAPTER
3 Carbohydrates and Lipids

John W. Baynes

LEARNING OBJECTIVES

After reading this chapter, you should be able to:

- Describe the structure and nomenclature of carbohydrates.
- Identify the major carbohydrates in our bodies and in our diet.
- Distinguish between reducing and nonreducing sugars.
- Describe various types of glycosidic bonds in oligosaccharides and polysaccharides.
- Identify the major classes of lipids in our bodies and in our diet.
- Describe the types of bonds in lipids and their sensitivity to saponification.
- Outline the general features of the fluid mosaic model of the structure of biological membranes.

INTRODUCTION

Carbohydrates and lipids are major sources of energy and are stored in the body as glycogen and triglycerides (fat)

This short chapter, which is largely an overview of collegiate studies, describes the structure of carbohydrates and lipids found in the diet and in tissues. These two classes of compounds differ significantly in physical and chemical properties. Carbohydrates are hydrophilic; the smaller carbohydrates (sugars) are soluble in aqueous solution, whereas large polymers such as starch or cellulose form colloidal dispersions or are insoluble. Lipids vary in size but rarely exceed 2 kDa in molecular mass; they are insoluble in water but soluble in organic solvents. Both carbohydrates and lipids may be bound covalently or noncovalently to proteins (glycoproteins, glycolipids, lipoproteins) and have important structural and regulatory functions, which are elaborated in later chapters. This chapter ends with a description of the **fluid mosaic model** of biological membranes, which describes how proteins, carbohydrates, and lipids are integrated into the structure of biological membranes that surround the cell and compartmentalize cellular functions.

CARBOHYDRATES

Nomenclature and structure of simple sugars

The classic definition of a carbohydrate is a polyhydroxy aldehyde, or ketone

The simplest carbohydrates, having two hydroxyl groups, are glyceraldehyde and dihydroxyacetone (Fig. 3.1). These three-carbon sugars are trioses; the suffix *-ose* designates a sugar. Glyceraldehyde is an **aldose**, and dihydroxyacetone is a **ketose** sugar. Prefixes and examples of longer-chain sugars are shown in Table 3.1.

Numbering of the carbons begins from the end containing the aldehyde, or ketone, functional group. Sugars are classified into the D or L family based on the configuration around the highest numbered asymmetric center (Fig. 3.2). In contrast to the L-amino acids, nearly all sugars found in the body have the D configuration.

An aldohexose, such as glucose, contains four asymmetric centers, so that there are 16 (2^4) possible stereoisomers, depending on whether each of the four carbons has the D or L configuration (Fig. 3.2). Eight of these aldohexoses are D-sugars (Fig. 3.2). There are four possible D-ketohexoses; fructose (fruit sugar; Fig. 3.2) is the only ketohexose present in significant concentration in our diet or in the body.

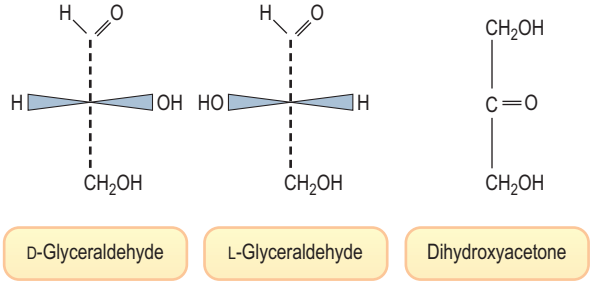


Fig. 3.1 **Structures of the trioses: D- and L-glyceraldehyde (aldoses) and dihydroxyacetone (a ketose).**

Table 3.1 Classification of carbohydrates by length of the carbon chain

Number of carbons	Name	Examples in human biology
Three	Triose	Glyceraldehyde, dihydroxyacetone
Four	Tetrose	Erythrose
Five	Pentose	Ribose, ribulose,* xylose, xylulose,* deoxyribose
Six	Hexose	Glucose, mannose, galactose, fucose, fructose
Seven	Heptose	Sedoheptulose*
Eight	Octose	None
Nine	Nonose	Neuraminic (sialic) acid

**The syllable* ul *indicates that the sugar is a ketose; the formal name for fructose would be* gluculose. *As with fructose (Fig. 3.2), the keto group is located at C-2 of the sugar, and the remaining carbons have the same geometry as the parent aldose sugar.*

Cyclization of sugars

Except for the trioses, sugars exist primarily in cyclic conformations. The linear sugar structures shown in Fig. 3.2 imply that aldose sugars have a chemically reactive, easily oxidizable, electrophilic, aldehyde residue. Aldehydes such as formaldehyde or glutaraldehyde react rapidly with amino groups in protein to form Schiff base (imine) adducts and crosslinks during fixation of tissues. However, glucose is relatively resistant to oxidation and does not react rapidly with protein. As shown in Fig. 3.3, glucose exists largely in nonreactive, inert, cyclic hemiacetal conformations, 99.99% in aqueous solution at pH 7.4 and 37°C. Of all the D-sugars in the world, D-glucose exists to the greatest extent in these cyclic conformations, making it the least oxidizable and least reactive with protein. It has been proposed that the relative chemical inertness of glucose is the reason for its evolutionary selection as blood sugar.

When glucose cyclizes to a hemiacetal, it may form a **furanose** or **pyranose** ring structure, named after the five- and six-carbon cyclic ethers, furan and pyran (Fig. 3.3). Note that the cyclization reaction creates a new asymmetric center at C-1, which is known as the **anomeric carbon**. The preferred conformation for glucose is the β-anomer (~65%), in which the hydroxyl group on C-1 is oriented equatorial to the ring. The β-anomer is the most stable form of glucose because all of the hydroxyl groups, which are bulkier than hydrogen, are oriented equatorially in the plane of the ring (Fig. 3.3), minimizing steric interactions. The α- and β-anomers of glucose can be isolated in pure form by selective crystallization from aqueous and organic solvents. They have different optical rotations but equilibrate over a period of hours in aqueous solution to form the equilibrium mixture of 65: 35 β:α anomers. These differences in structure may seem unimportant, but, in fact, metabolic pathways generally use one anomer but not the other. Similarly, although the fructopyranose conformations are the primary forms of fructose in aqueous solution, fructose metabolism proceeds from furanose conformations.

In addition to the basic sugar structures just discussed, the structures of other common sugars are shown in Fig. 3.4. These sugars—deoxysugars, aminosugars, and sugar acids—are found primarily in oligomeric or polymeric structures in the body (e.g., ribose in RNA and deoxyribose in DNA), or they may be attached to proteins or lipids to form glycoconjugates (glycoproteins or glycolipids, respectively). **Glucose (blood sugar) is the only sugar found to a significant extent as a free sugar in the body**.

Disaccharides, oligosaccharides, and polysaccharides

Sugars are linked to one another by glycosidic bonds to form complex glycans

Carbohydrates are linked to one another by glycosidic bonds to form disaccharides, trisaccharides, oligosaccharides, and polysaccharides. Polysaccharides composed of a single sugar are termed *homoglycans*, whereas those with complex compositions are termed *heteroglycans*. The name of the more complex structures includes not only the name of the component sugars but also the ring conformation of the sugars, the anomeric configuration of the linkage between sugars, the site of attachment of one sugar to another, and the nature of the atom involved in the linkage - usually an oxygen or *O*-glycosidic bond, sometimes a nitrogen or *N*-glycosidic bond. Fig. 3.5 shows the structure of several common disaccharides in our diet: **lactose** (milk sugar); **sucrose** (table sugar); **maltose** and isomaltase (products of digestion of starch); cellobiose, which is obtained on hydrolysis of **cellulose;** and **hyaluronic acid**.

Differences in linkage of sugars make a big difference in metabolism and nutrition

Amylose, a component of **starch**, is an α-1→4-linked linear glucan (see maltose in Fig. 3.5 and glycogen in Fig. 12.2), whereas **cellulose** (Fig. 3.5) is a β-1→4-linked linear glucan. These two polysaccharides differ only in the anomeric linkage between glucose subunits, but they are very different molecules. Starch forms a colloidal suspension in water, whereas cellulose is insoluble; starch is pasty, whereas cellulose is fibrous; starch is digestible, whereas cellulose is indigestible by humans; starch is a food, rich in calories, whereas cellulose is roughage.

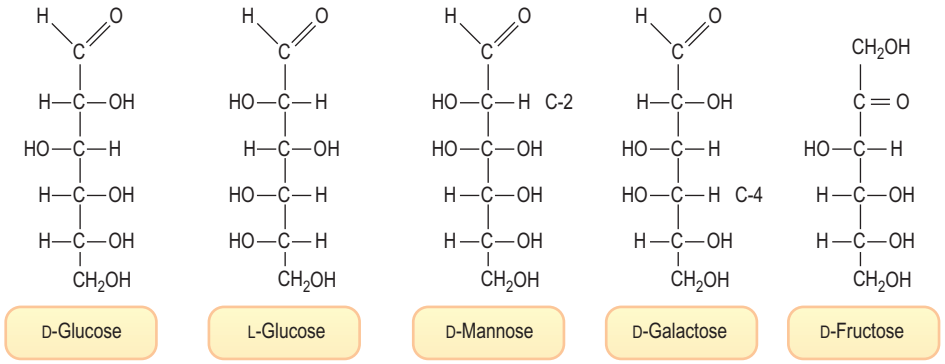


Fig. 3.2 **Structures of hexoses: D- and L-glucose, D-mannose, D-galactose, and D-fructose.** These linear projections of carbohydrate structures are known as Fischer projections. The D and L designations are based on the configuration at the highest numbered asymmetric center, C-5 in the case of hexoses. Note that L-glucose is the mirror image of D-glucose (i.e., the geometry at all of the asymmetric centers is reversed). Mannose is the C-2 epimer and galactose the C-4 epimer of glucose; epimers differ at only one stereoisomeric center.

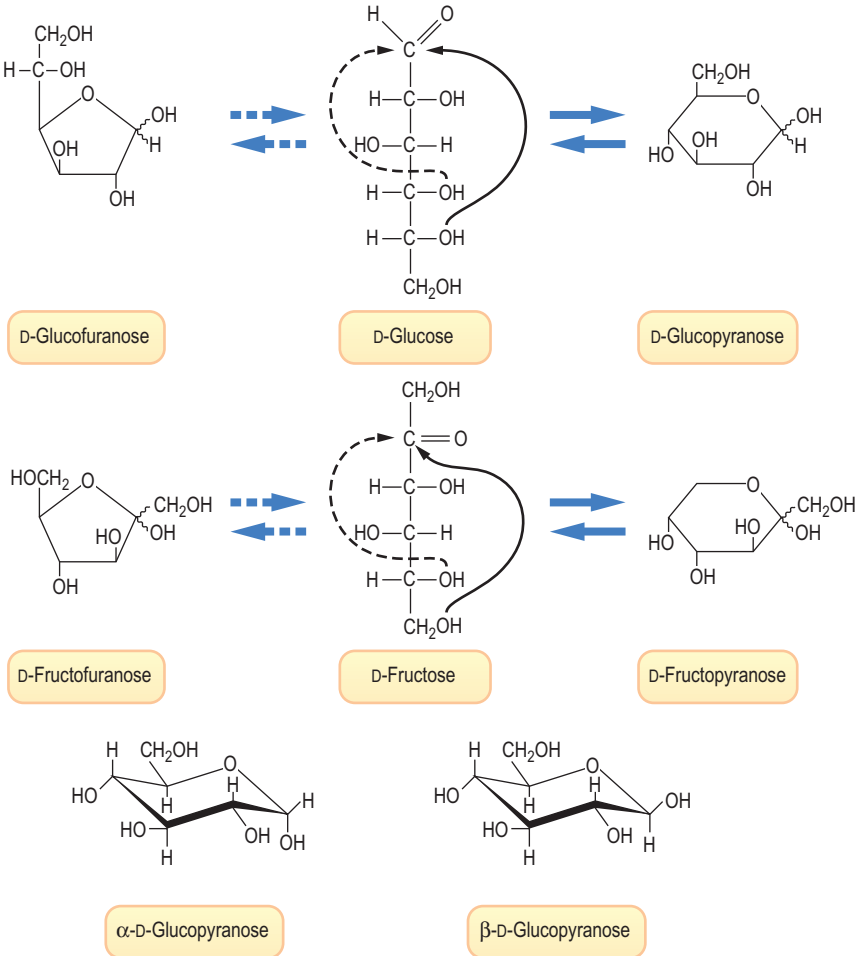


Fig. 3.3 **Linear and cyclic representations of glucose and fructose.** (Top) There are four cyclic conformations of glucose, in equilibrium with the linear form: α- and β-glucopyranose and α- and β-glucofuranose. The pyranose forms account for more than 99% of total glucose in solution. These cyclic structures are known as Haworth projections; by convention, groups to the right in Fischer projections are shown below the ring, and groups to the left are shown above the ring. The squiggly bonds to H and OH from C-1, the anomeric carbon, indicate indeterminate geometry and represent either the α or the β anomer. (Middle) The linear and cyclic forms of fructose. The ratio of pyranose:furanose forms of fructose in aqueous solution is ~3: 1. The ratio shifts as a function of temperature, pH, salt concentration, and other factors. (Bottom) Stereochemical representations of the chair forms of α- and β-glucopyranose. The preferred structure in solution, β-glucopyranose, has all of the hydroxyl groups, including the anomeric hydroxyl group, in equatorial positions around the ring, minimizing steric interactions.

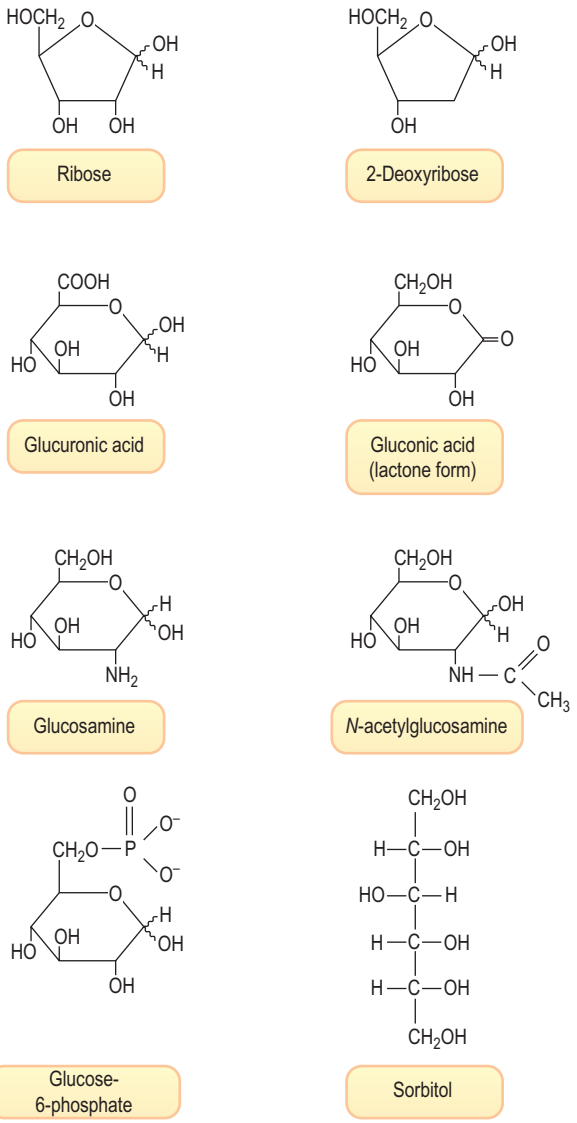


Fig. 3.4 **Examples of various types of sugars found in human tissues.** Ribose, the pentose sugar in ribonucleic acid (RNA); 2-deoxyribose, the deoxypentose in DNA; glucuronic acid, an acidic sugar formed by oxidation of C-6 of glucose; gluconic acid, an acidic sugar formed by oxidation of C-1 of glucose, shown in the δ-lactone form; glucosamine, an amino sugar; *N*-acetylglucosamine, an acetylated amino sugar; glucose-6-phosphate, a phosphate ester of glucose, an intermediate in glucose metabolism; sorbitol, a polyol formed on reduction of glucose.

LIPIDS

Lipids are found primarily in three compartments in the body: plasma, adipose tissue, and biological membranes

This introduction focuses on the structure of **fatty acids** (the simplest form of lipids, found primarily in plasma), **triglycerides** (the storage form of lipids, found primarily in adipose tissue), and **phospholipids** (the major class of membrane lipids in all cells). Steroids, such as cholesterol, and (glyco) sphingolipids are mentioned in the context of biological membranes, but these lipids and others, such as plasmalogens, polyisoprenoids, and eicosanoids, are addressed in detail in later chapters.

ADVANCED CONCEPT BOX
THE INFORMATION CONTENT OF COMPLEX GLYCANS

Sugars are attached to each other in **glycosidic linkages** between a hemiacetal carbon of one sugar and a hydroxyl group of another sugar. Two glucose residues can be linked in many different linkages (i.e., α1,2; α1,3; α1,4; α1,6; β1,2; β1,3; β1,4; β1,6; α,α1,1; α,β1,1; β,β1,1) to give 11 different disaccharides, each with different chemical and biological properties. Two different sugars, such as glucose and galactose, can be linked, either glucose → galactose or galactose → glucose, and these two disaccharides can have a total of 20 different isomers.

In contrast, two identical amino acids, such as two alanines, can form only one dipeptide, alanyl-alanine. And two different amino acids (e.g., alanine and glycine) can form only two dipeptides (e.g., alanyl-glycine and glycyl-alanine). As a result, sugars have the potential to provide a great deal of chemical information. As outlined in Chapters 17–19, carbohydrates bound to proteins and lipids in cell membranes can serve as recognition signals for both cell–cell and cell–pathogen interactions.

CLINICAL TEST BOX
REDUCING SUGAR ASSAY FOR BLOOD GLUCOSE

The original assays for blood glucose measured the reducing activity of blood. These assays work because glucose, at 5-mM concentration, is the major reducing substance in the blood. The Fehling and Benedict assays use alkaline cupric salt solutions. With heating, the glucose decomposes oxidatively, yielding a complex mixture of organic acids, aldehydes, ketones, and enol and dicarbonyl compounds. Oxidation of these carbonyl compounds reduces cupric ion (blue-green color) to cuprous ion (orange-red color) in solution. The color yield produced is directly proportional to the glucose content of the sample.

Reducing sugar assays do not distinguish glucose from other reducing sugars, such as fructose or galactose. In diseases of fructose and galactose metabolism, such as hereditary fructose intolerance or galactosemia (Chapter 17), these assays could yield positive results for high blood sugar, creating the false impression of diabetes. Sucrose and gluconic acid (Figs. 3.4 and 3.5) are nonreducing sugars, lacking an aldehyde or keto group, and yield a negative reaction in reducing sugar assays. Starch and cellulose are also considered non-reducing carbohydrates because the reducing ends are such a small fraction of the total sugar groups in the polysaccharide chains.

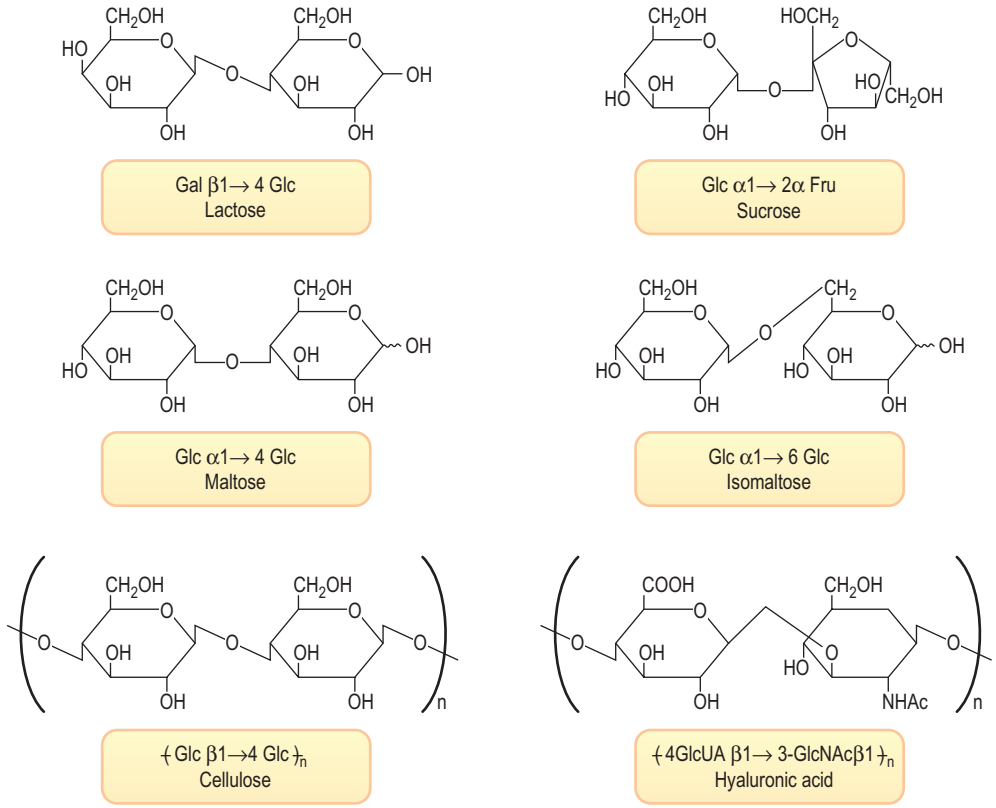


Fig. 3.5 **Structures of common disaccharides and polysaccharides.** Lactose (milk sugar); sucrose (table sugar); maltose and isomaltose, disaccharides formed on degradation of starch; and repeating disaccharide units of cellulose (from wood) and hyaluronic acid (from vertebral disks). Glycogen (not shown; see Fig. 12.2) is an α-1,4 linked glucan with α-1-6 branches. Fru, fructose; Gal, galactose; Glc, glucose; GlcNAc, *N*-acetylglucosamine; GlcUA, glucuronic acid.

Fatty acids

Fatty acids exist in free form and as components of more complex lipids

As summarized in Table 3.2, most fatty acids in the body are long, straight-chain alkanoic acids, with 16 or 18 carbons. They may be saturated or unsaturated, the latter containing up to five double bonds, all in *cis* geometry. The double bonds are not conjugated but separated by methylene groups; solid lipids are generally white or opalescent in color.

Fatty acids with a single double bond are described as monounsaturated, and those with two or more double bonds are described as polyunsaturated fatty acids. The polyunsaturated fatty acids are commonly classified into two groups, **ω-3 and ω-6 fatty acids**, depending on whether the first double bond appears three or six carbons from the terminal methyl group. The melting point of fatty acids, as well as that of more complex lipids, increases with the chain length of the fatty acid but decreases with the number of double bonds. The ***cis*-double bonds** place a kink in the linear structure of the fatty acid chain, interfering with close packing, therefore requiring a lower temperature for freezing (i.e., they have a lower melting point).

Triacylglycerols (triglycerides)

Triglycerides are the storage form of lipids in adipose tissue

Fatty acids in plant and animal tissues are commonly esterified to glycerol, forming a triacylglycerol (triglyceride; Fig. 3.6), either oils (liquid) or fats (solid). In humans, triglycerides are stored in solid form in adipose tissue as fat. They are degraded to glycerol and fatty acids in response to hormonal signals, then released into plasma for metabolism in other tissues, primarily muscle and liver. The ester bond of triglycerides and other glycerolipids is also readily hydrolyzed in vitro by a strong base, such as NaOH, forming glycerol and free fatty acids. This process is known as **saponification;** one of the products, the sodium salt of the fatty acid, is soap.