I'm not robot
reCAPTCHA
Continue

Essential biochemistry 3rd edition pdf textbook online

Loading PreviewSorry, preview is currently unavailable. You can download the paper by clicking the button above. Delivery Update: We're currently shipping orders daily. However, due to supply issues for print and paper in some geographies, deliveries may be delayed. To provide our customers with immediate access to content, we are offering a 50% discount on Science & Technology print & eBook bundles for instant downloading of eBooks: Terms & conditionsEssentials of Medical Biochemistry, Third Edition offers a condensed, yet detailed overview of clinical biochemistry, spanning fundamentals and relevant physiologic and pathophysiologic concepts. Pivotal clinical case studies aid in understanding basic science in the context of diagnosis and treatment of human diseases, and the text illuminates key topics in molecular immunology and hemostasis. Users will find fundamental concepts aiding students and professionals in biochemistry, medicine, and other healthcare disciplines. The text is a useful refresher that will help users meet USMLE and other professional licensing examination requirements, providing thorough introductions, key points, multicolored illustrations of chemical structures and figures, fact-filled tables, and recommended reading lists. This Third Edition has been fully updated to address evolving techniques in the biological sciences, including genomics, metabolomics, transcriptomics, epigenomics, proteomics, and gene therapy, among other methods. In addition, each chapter has been fully revised for current science and now features learning objectives and chapter summaries, supplemental reading, and 5 clinical case based multiple choice questions. New clinical cases have been added throughout. Integrates the biochemical principles with physiological, pharmacological, and pathological aspects of human diseases Each chapter features learning objectives, summaries, required and supplemental reading lists, clinical cases, and multiple-choice questions Presents essential biochemical concepts within the context of their biological functions Offers instructional overview figures, flowcharts, tables and multi-colored illustrations Provides an online ancillary package with PowerPoint images and an additional 500 study questions to aid in comprehension and USMLE exam preparationMedical students, dentistry students, pharmaceutical students, and graduate students in medical biochemistry; Active researchers in biochemistry, molecular biology, pharmaceutical science, and cellular biology; medical chemists; Clinicians1. The Human Organism: Organ Systems, Cells, Organelles, and Our Microbiota2. Water, Acids, Bases, and Buffers3. Amino Acids4. Three-Dimensional Structure of Proteins and Disorders of Protein Misfolding5. Energetics of Biological Systems, Enzymes and Enzyme Regulation6. Clinical Enzymology and Biomarkers of Tissue Injury7. Simple Carbohydrates8. Heteropolysaccharides: Glycoconjugates, Glycoproteins, and Glycolipids9. Connective Tissue: Fibrous and Nonfibrous Proteins and Proteoglycans10. Gastrointestinal Digestion and Absorption11. Carbohydrate Metabolism I: Glycolysis and the Tricarboxylic Acid Cycle12. Electron Transport Chain, Oxidative Phosphorylation, and Other Oxygen-Consuming Systems13. Carbohydrate Metabolism II: Gluconeogenesis, Glycogen Synthesis and Breakdown, and Alternative Pathways14. Protein and Amino Acid Metabolism15. Lipids I: Fatty Acids and Eicosanoids16. Lipids II: Phospholipids, Glycosphingolipids, and Cholesterol17. Lipids III: Plasma Lipoproteins18. Contractile Systems19. Perturbations of Energy Metabolism: Obesity and Diabetes Mellitus20. Structure and Properties of DNA21. DNA Replication, Repair, and Mutagenesis22. RNA and Protein Synthesis23. Regulation of Gene Expression24. Nucleotide Metabolism25. Hemoglobin and Metabolism of Iron & Heme26. Endocrine Metabolism I: Introduction and Signal Transduction27. Endocrine Metabolism II: Hypothalamus and Pituitary28. Endocrine Metabolism III: Adrenal Glands29. Endocrine Metabolism IV: Thyroid Gland30. Endocrine Metabolism V: Reproductive System31. Immunology32. Biochemistry of Hemostasis33. Mineral Metabolism34. Vitamin Metabolism35. Drug MetabolismNo. of pages: 840Language: EnglishCopyright: © Academic Press 2022Published: August 1, 2022Imprint: Academic PressPaperback ISBN: 9780323885416Chung-Eun Ha is an Associate Professor in the department of Native Hawaiian Healthat the John A. Burns School of Medicine, University of Hawaii at Manoa. Since 2001,he has been teaching medical biochemistry to students who come from disadvantagedbackgrounds. He is also actively involved in the various aspects of molecular biologyand biochemical research in the field of clinical medicine.Associate Professor, Department of Native Hawaiian Health, John A. Burns School of Medicine, University of Hawaii at Manoa, Honolulu, HI, USAN. V. Bhagavan is an Emeritus Professor in the department of Anatomy, Biochemistry,and Physiology at the John A. Burns School of Medicine, University of Hawaii at Manoa.An active professor, researcher, and teacher, he has many years of experience in boththe instruction and practice of clinical biochemistry.Emeritus Professor, Department of Anatomy, Biochemistry, and Physiology, John A. Burns School of Medicine, University of Hawaii at Manoa, Honolulu, HI, USAWrite a reviewThere are currently no reviews for "Essentials of Medical Biochemistry" Add a review and share your thoughts with other readers. Be the first. Add a review and share your thoughts with other readers. Be the first. E S S E N T I A L B I O C H E M I S T RY THIRD EDITION CHARLOTTE W. PRATT KATHLEEN CORNELY This page is intentionally left blank USEFUL CONSTANTS Avogadro's number Gas constant (R) Faraday (F) Kelvin (K) 6.02 3 1023 molecules ? mol21 8.314 J ? K21 ? mol21 96,485 J ? V21 ? mol21 8C 1 273 KEY EQUATIONS Henderson–Hasselbalch equation pH 5 pK 1 log [A2] [HA] Michaelis–Menten equation v0 5 Vmax [S] KM 1 [S] Lineweaver–Burk equation KM 1 1 1 1 5a b v0 V max [S] V max Nernst equation E 5 E8¿ 2 0.026 V [A reduced] RT [A reduced] ln or E 5 E8¿ 2 ln n nF [A oxidized] [A oxidized] Thermodynamics equations DG 5 DH 2 T D S DG 8¿ 5 2RT ln K eq DG 5 DG 8¿ 1 RT ln DG 8¿ 5 2nFDE8¿ [C][D] [A][B] Common Functional Groups and Linkages in Biochemistry Structurea Compound Name Amineb RNH2 or R2NH or R3N or Functional Group RNH⩽ 3 R2NH⩽ 2 R3NH⩽ N ⩽ or N (amino group) Alcohol ROH OH(hydroxyl group) Thiol RSH SH (sulfhydryl group) Ether ROR O Aldehyde Ketone R R O O C H C O O C R (carbonyl group), R C O (carbonyl group), R C R C O OH or C O OH(carboxyl group) or R C O⩾ C O⩾ (carboxylate group) R C (acyl group) (acyl group) O O O Ester O C O Carboxylic acidb (Carboxylate) (ether linkage) C OR O (ester linkage) C N (amido group) C or O Amide Imineb R C O NH2 R C O NHR R C NR2 R R NH or NR or O R R NH⩽ 2 NHR⩽ O Phosphoric acid esterb R O P O P OH or O O P O⩾ O P O⩾ a P O OH or P O O P O OH or P O⩾ P O O OH OH O O O P P O O (phosphoanhydride linkage) OH O O O⩾ O⩾ (phosphoryl group, Pi) O⩾ P O OH or P O O P O⩾ O⩾ OH O⩾ OH (diphosphoryl group, pyrophosphoryl group, PPi) R represents any carbon-containing group. In a molecule with more than one R group, the groups may be the same or different. Under physiological conditions, these groups are ionized and hence bear a positive or negative charge. b H (imino group) (phosphoester linkage) O OH OH O R ⩽ N OH O R P O O⩾ Diphosphoric acid esterb C O OH O R N Third Edition Essential Biochemistry This page is intentionally left blank Third Edition Essential Biochemistry CHARLOTTE W. PRATT Seattle Pacific University KATHLEEN CORNELY Providence College VICE PRESIDENT & EXECUTIVE PUBLISHER VICE PRESIDENT, PUBLISHER SPONSORING EDITOR EDITORIAL ASSISTANT SENIOR MARKETING MANAGER MARKETING ASSISTANT CONTENT MANAGER PRODUCTION EDITOR CONTENT EDITOR PHOTO DEPARTMENT MANAGER PHOTO RESEARCHER DESIGN DIRECTOR SENIOR DESIGNER COVER DESIGNER PRODUCTION MANAGEMENT SERVICES Kaye Pace Petra Recter Joan Kalkut Susan Tsui Kristine Ruff Andrew Ginsberg Juanita Thompson Sandra Dumas Alyson Rentrop Hilary Newman Teri Stratford Harry Nolan Wendy Lai Thomas Nery Suzanne Ingrao Cover Photo: ©Volker Steger/Science Source/Photo Researchers, Inc. This book was typeset in 10.5/12 Adobe Garamond at Aptara and printed and bound by Courier/Kendallville. The cover was printed by Courier/Kendallville. Founded in 1807, John Wiley & Sons, Inc. has been a valued source of knowledge and understanding for more than 200 years, helping people around the world meet their needs and fulfill their aspirations. Our company is built on a foundation of principles that include responsibility to the communities we serve and where we live and work. In 2008, we launched a Corporate Citizenship Initiative, a global effort to address the environmental, social, economic, and ethical challenges we face in our business. Among the issues we are addressing are carbon impact, paper specifications and procurement, ethical conduct within our business and among our vendors, and community and charitable support. For more information, please visit our website: www.wiley.com/go/citizenship. The paper in this book was manufactured by a mill whose forest management programs include sustained yieldharvesting of its timberlands. Sustained yield harvesting principles ensure that the number of trees cut each year does not exceed the amount of new growth. This book is printed on acid-free paper. Copyright © 2014, 2011, 2004 John Wiley and Sons, Inc.. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 646-8600. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030-5774, (201) 748-6011, fax (201) 748-6008. Evaluation copies are provided to qualified academics and professionals for review purposes only, for use in their courses during the next academic year. These copies are licensed and may not be sold or transferred to a third party. Upon completion of the review period, please return the evaluation copy to Wiley. Return instructions and a free of charge return shipping label are available at www.wiley.com/go/returnlabel. Outside of the United States, please contact your local representative. ISBN 13 978-1118-08350-5 ISBN 13 978-1118-44168-8 Printed in the United States of America. 10 9 8 7 6 5 4 3 2 1 ABOUT THE AUTHORS Charlotte Pratt received a B.S. in biology from the University of Notre Dame and a Ph.D. in biochemistry from Duke University. She is a protein chemist who has conducted research in blood coagulation and inflammation at the University of North Carolina at Chapel Hill. She is currently Associate Professor in the Biology Department at Seattle Pacific University. Her interests include molecular evolution, enzyme action, and the relationship between metabolic processes and disease. She has written numerous research and review articles, has worked as a textbook editor, and is a co-author, with Donald Voet and Judith G. Voet, of Fundamentals of Biochemistry, published by John Wiley & Sons, Inc. Kathleen Cornely holds a B.S. in chemistry from Bowling Green (Ohio) State University, an M.S. in biochemistry from Indiana University, and a Ph.D. in nutritional biochemistry from Cornell University. Her experimental research has included a wide range of studies of protein purification and chemical modification. She currently serves as a Professor of Chemistry and Biochemistry at Providence College where she has taught courses in biochemistry, organic chemistry, and general chemistry. Her recent research efforts have focused on chemical education, particularly the use of case studies and guided inquiry in biochemistry education. She serves on the editorial board of Biochemistry and Molecular Biology Education and is a member of the American Society for Biochemistry and Molecular Biology. v BRIEF CONTENTS part one FOUNDATIONS 1 THE CHEMICAL BASIS OF LIFE 2 AQUEOUS CHEMISTRY part two 1 24 MOLECULAR STRUCTURE AND FUNCTION 3 FROM GENES TO PROTEINS 4 PROTEIN STRUCTURE 5 PROTEIN FUNCTION 51 87 121 6 HOW ENZYMES WORK 158 7 ENZYME KINETICS AND INHIBITION 8 LIPIDS AND MEMBRANES 220 9 MEMBRANE TRANSPORT 10 SIGNALING 241 266 11 CARBOHYDRATES part three 290 METABOLISM 12 METABOLISM AND BIOENERGETICS 13 GLUCOSE METABOLISM 14 THE CITRIC ACID CYCLE 338 370 15 OXIDATIVE PHOSPHORYLATION 16 PHOTOSYNTHESIS 394 420 17 LIPID METABOLISM 442 18 NITROGEN METABOLISM 475 19 REGULATION OF MAMMALIAN FUEL METABOLISM part four 509 GENETIC INFORMATION 20 DNA REPLICATION AND REPAIR 21 TRANSCRIPTION AND RNA 22 PROTEIN SYNTHESIS vi 188 590 561 529 308 CONTENTS part one FOUNDATIONS 1 THE CHEMICAL BASIS OF LIFE 1-1 1-2 What is Biochemistry? 2 Biological Molecules 3 Cells contain four major types of biomolecules 5 There are three major kinds of biological polymers 6 Energy and Metabolism 10 Enthalpy and entropy are components of free energy 10 ©G is less than zero for a spontaneous process 11 Life is thermodynamically possible 12 The Origin and Evolution of Life 14 The prebiotic world 14 Origins of modern cells 16 1-3 1-4 1 BOX 1-A BIOCHEMISTRY NOTE Units Used in Biochemistry 7 2-2 2-3 The Hydrophobic Effect 30 Amphiphilic molecules experience both hydrophilic interactions and the hydrophobic effect 31 The hydrophobic core of a lipid bilayer is a barrier to diffusion 32 Acid–Base Chemistry 33 [H⩽] and [OH⩾] are inversely related 34 The pH of a solution can be altered 36 A pK value describes an acid's tendency to ionize 37 The pH of a solution of acid is related to the pK 37 2-4 Tools and Techniques: Buffers 41 BOX 2-A BIOCHEMISTRY NOTE Why Do Some Drugs Contain Fluorine? 28 BOX 1-B BIOCHEMISTRY NOTE How Does Evolution Work? 17 BOX 2-B BIOCHEMISTRY NOTE Sweat, Exercise, and Sports Drinks 33 BIOINFORMATICS PROJECT 1 The Biochemical Literature 19 BOX 2-C BIOCHEMISTRY NOTE Atmospheric CO2 and Ocean Acidification 35 2 AQUEOUS CHEMISTRY BOX 2-D CLINICAL CONNECTION Acid–Base Balance in Humans 43 2-1 Water Molecules form Hydrogen Bonds 25 Hydrogen bonds are one type of electrostatic force 26 Water dissolves many compounds 29 part two 24 MOLECULAR STRUCTURE AND FUNCTION 3 FROM GENES TO PROTEINS 3-1 DNA Is the Genetic Material 52 Nucleic acids are polymers of nucleotides 53 Some nucleotides have other functions 54 DNA is a double helix 56 RNA is single-stranded 59 DNA can be denatured and renatured 59 51 3-2 Genes Encode Proteins 61 A mutated gene can cause disease 63 3-3 Genomics 65 Gene number is roughly correlated with organismal complexity 65 Genes are identified by comparing sequences 67 Genomic data can be linked to disease 68 3-4 Tools and Techniques: Manipulating DNA 70 DNA sequencing uses DNA polymerase to make a complementary strand 70 The polymerase chain reaction amplifies DNA 72 Restriction enzymes cut DNA at specific sequences 75 DNA fragments are joined to produce recombinant DNA 76 Cloned genes yield valuable products 77 Genetically modified organisms have practical applications 79 Gene therapy can cure some human diseases 79 BOX 3-A CLINICAL CONNECTION Discovery of the Cystic Fibrosis Gene 64 vii viii CONTENTS BOX 3-B BIOCHEMISTRY NOTE DNA Fingerprinting 74 BIOINFORMATICS PROJECT 2 Databases for the Storage and "Mining" of Genome Sequences 81 4 PROTEIN STRUCTURE 87 4-1 Proteins Are Chains of Amino Acids 89 The 20 amino acids have different chemical properties 89 Peptide bonds link amino acids in proteins 91 The amino acid sequence is the first level of protein structure 95 4-2 Secondary Structure: The Conformation of the Peptide Group 96 The ␣ helix exhibits a twisted backbone conformation 97 The ␣ sheet contains multiple polypeptide strands 97 Proteins also contain irregular secondary structure 98 4-3 Tertiary Structure and Protein Stability 99 Proteins have hydrophobic cores 100 Protein structures are stabilized mainly by the hydrophobic effect 102 Cross-links help stabilize proteins 102 Protein folding begins with the formation of secondary structures 104 4-4 Quaternary Structure 107 4-5 Tools and Techniques: Analyzing Protein Structure 108 Chromatography takes advantage of a polypeptide's unique properties 108 Mass spectrometry reveals amino acid sequences 111 Protein structures are determined by X-ray crystallography, electron crystallography, and NMR spectroscopy 113 BOX 4-A BIOCHEMISTRY NOTE Does Chirality Matter? 89 BOX 4-B BIOCHEMISTRY NOTE Monosodium Glutamate 92 BOX 4-C CLINICAL CONNECTION Protein Misfolding and Disease 105 BOX 4-D BIOCHEMISTRY NOTE Mass Spectrometry Applications 113 BIOINFORMATICS PROJECT 3 Visualizing ThreeDimensional Protein Structures 115 5 PROTEIN FUNCTION 5-1 Myoglobin and Hemoglobin: Oxygen-Binding Proteins 122 Oxygen binding to myoglobin depends on the oxygen concentration 123 121 Myoglobin and hemoglobin are related by evolution 124 Oxygen binds cooperatively to hemoglobin 126 A conformational shift explains hemoglobin's cooperative behavior 128 H⩽ ions and bisphosphoglycerate regulate oxygen binding to hemoglobin in vivo 129 5-2 Structural Proteins 133 Microfilaments are made of actin 133 Actin filaments continuously extend and retract 134 Tubulin forms hollow microtubules 136 Some drugs affect microtubules 137 Keratin is an intermediate filament 138 Collagen is a triple helix 140 Collagen molecules are covalently cross-linked 142 5-3 Motor Proteins 144 Myosin has two heads and a long tail 144 Myosin operates through a lever mechanism 146 Kinesin is a microtubule-associated motor protein 146 Kinesin is a processive motor 150 BOX 5-A BIOCHEMISTRY NOTE Erythropoietin Boosts Red Blood Cell Production 126 BOX 5-B BIOCHEMISTRY NOTE Carbon Monoxide Poisoning 127 BOX 5-C CLINICAL CONNECTION Hemoglobin Mutations 130 BOX 5-D BIOCHEMISTRY NOTE Vitamin C Deficiency Causes Scurvy 141 BOX 5-E CLINICAL CONNECTION Genetic Collagen Diseases 143 BOX 5-F BIOCHEMISTRY NOTE Myosin Mutations and Deafness 147 BIOINFORMATICS PROJECT 4 Using Databases to Compare and Identify Related Protein Sequences 6 6-1 HOW ENZYMES WORK 151 158 What Is an Enzyme? 159 Enzymes are usually named after the reaction they catalyze 161 6-2 The Chemistry of Catalysis 162 A catalyst provides a reaction pathway with a lower activation energy barrier 164 Enzymes use chemical catalytic mechanisms 164 The catalytic triad of chymotrypsin promotes peptide bond hydrolysis 169 6-3 The Unique Properties of Enzyme Catalysts 171 Enzymes stabilize the transition state 171 Efficient catalysis depends on proximity and orientation effects 173 The active site microenvironment promotes catalysis 173 CONTENTS 6-4 Some Additional Features of Enzymes 174 Not all serine proteases are related by evolution 174 Enzymes with similar mechanisms exhibit different substrate specificity 175 Chymotrypsin is activated by proteolysis 176 Protease inhibitors limit protease activity 179 BOX 6-A BIOCHEMISTRY NOTE Depicting Reaction Mechanisms 167 BOX 6-B CLINICAL CONNECTION Blood Coagulation Requires a Cascade of Proteases 177 7 ENZYME KINETICS AND INHIBITION 188 7-1 Introduction to Enzyme Kinetics 189 7-2 Derivation and Meaning of the 7-3 Michaelis–Menten Equation 191 Rate equations describe chemical processes 191 The Michaelis–Menten equation is a rate equation for an enzyme-catalyzed reaction 192 KM is the substrate concentration at which velocity is half-maximal 195 The catalytic constant describes how quickly an enzyme can act 195 kcat/KM indicates catalytic efficiency 195 KM and Vmax are experimentally determined 196 Not all enzymes fit the simple Michaelis–Menten model 198 Enzyme Inhibition 200 Some inhibitors act irreversibly 202 Competitive inhibition is the most common form of reversible enzyme inhibition 203 Transition state analogs inhibit enzymes 205 Other types of inhibitors affect Vmax 207 Allosteric enzyme regulation includes inhibition and activation 209 Several factors may influence enzyme activity 211 BOX 7-A CLINICAL CONNECTION Drug Development 200 8-3 Membrane Proteins 230 Integral membrane proteins span the bilayer 231 An ␣ helix can cross the bilayer 231 A transmembrane ␣ sheet forms a barrel 232 Lipid-linked proteins are anchored in the membrane 232 8-4 The Fluid Mosaic Model 233 Membrane glycoproteins face the cell exterior 234 BOX 8-A BIOCHEMISTRY NOTE Omega-3 Fatty Acids 222 BOX 8-B CLINICAL CONNECTION The Lipid Vitamins A, D, E, and K 225 9 MEMBRANE TRANSPORT 241 9-1 The Thermodynamics of Membrane Transport 242 Ion movements alter membrane potential 243 Transporters mediate transmembrane ion movement 245 9-2 Passive Transport 246 Porins are ␣ barrel proteins 246 Ion channels are highly selective 247 Gated channels undergo conformational changes 248 Aquaporins are water-specific pores 250 Some transport proteins alternate between conformations 251 9-3 Active Transport 252 The Na,K-ATPase changes conformation as it pumps ions across the membrane 252 ABC transporters mediate drug resistance 253 Secondary active transport exploits existing gradients 254 9-4 Membrane Fusion 255 SNAREs link vesicle and plasma membranes 256 Membrane fusion requires changes in bilayer curvature 259 BOX 7-B BIOCHEMISTRY NOTE Inhibitors of HIV Protease 206 BOX 9-A BIOCHEMISTRY NOTE Pores Can Kill 248 BIOINFORMATICS PROJECT 5 Enzyme Inhibitors and Rational Drug Design 212 BOX 9-B BIOCHEMISTRY NOTE Some Drugs Interfere with Neuronal Signaling 257 8 8-1 LIPIDS AND MEMBRANES 220 Lipids 221 Fatty acids contain long hydrocarbon chains 221 Some lipids contain polar head groups 223 Lipids perform a variety of physiological functions 225 8-2 The Lipid Bilayer 227 The bilayer is a fluid structure 227 Natural bilayers are asymmetric 230 ix BOX 9-C CLINICAL CONNECTION Antidepressants Block Serotonin Transport 258 10 SIGNALING 266 10-1 General Features of Signaling Pathways 267 A ligand binds to a receptor with a characteristic affinity 267 Most signaling occurs through two types of receptors 269 The effects of signaling are limited 270 x CONTENTS 10-2 G Protein Signaling Pathways 271 G protein–coupled receptors include seven transmembrane helices 271 The receptor activates a G protein 272 Adenylate cyclase generates the second messenger cyclic AMP 273 Cyclic AMP activates protein kinase A 273 Signaling pathways are also switched off 275 The phosphoinositide signaling pathway generates two second messengers 276 Calmodulin mediates some Ca2⩽ signals 277 10-3 Receptor Tyrosine Kinases 277 The insulin receptor has two ligand-binding sites 277 The receptor undergoes autophosphorylation 278 10-4 Lipid Hormone Signaling 281 Eicosanoids are short-range signals 282 BOX 10-A BIOCHEMISTRY NOTE Bacterial Quorum Sensing 268 BOX 10-B BIOCHEMISTRY NOTE Cell Signaling and Cancer 280 290 11-1 Monosaccharides 291 Most carbohydrates are chiral compounds 291 Cyclization generates ␣ and ␤ anomers 292 Monosaccharides can be derivatized in many different ways 293 11-2 Polysaccharides 294 Lactose and sucrose are the most common disaccharides 295 Starch and glycogen are fuel storage molecules 296 Cellulose and chitin provide structural support 296 Bacterial polysaccharides form a biofilm 298 11-3 Glycoproteins 299 N-linked oligosaccharides undergo processing 299 O-linked oligosaccharides tend to be large 300 What is the purpose of the oligosaccharide groups? 300 Proteoglycans contain long glycosaminoglycan chains 301 Bacterial cell walls are made of peptidoglycan 302 297 BOX 11-B BIOCHEMISTRY NOTE The ABO Blood Group System 301 BOX 10-D BIOCHEMISTRY NOTE Aspirin and Other Inhibitors of Cyclooxygenase 283 12 CARBOHYDRATES BOX 11-A BIOCHEMISTRY NOTE Cellulosic Biofuel BOX 10-C BIOCHEMISTRY NOTE Oral Contraceptives 282 part three 11 BOX 11-C BIOCHEMISTRY NOTE Antibiotics and Bacterial Cell Walls 303 METABOLISM METABOLISM AND BIOENERGETICS 308 12-1 Food and Fuel 309 Cells take up the products of digestion 309 Monomers are stored as polymers 312 Fuels are mobilized as needed 312 12-2 Metabolic Pathways 314 Some major metabolic pathways share a few common intermediates 315 Many metabolic pathways include oxidation– reduction reactions 316 Metabolic pathways are complex 318 Human metabolism depends on vitamins 320 12-3 Free Energy Changes in Metabolic Reactions 323 The free energy change depends on reactant concentrations 323 Unfavorable reactions are coupled to favorable reactions 325 Free energy can take different forms 328 Regulation occurs at the steps with the largest free energy changes 330 BOX 12-A BIOCHEMISTRY NOTE Dietary Guidelines 311 BOX 12-B BIOCHEMISTRY NOTE The Transcriptome, the Proteome, and the Metabolome 319 BOX 12-C BIOCHEMISTRY NOTE What Is a Calorie? 324 BOX 12-D BIOCHEMISTRY NOTE Powering Human Muscles 328 BIOINFORMATICS PROJECT 6 Metabolic Enzymes, Microarrays, and Proteomics 332 13 GLUCOSE METABOLISM 338 13-1 Glycolysis 339 Reactions 1–5 are the energy-investment phase of glycolysis 341 Reactions 6–10 are the energy-payoff phase of glycolysis 345 Pyruvate is converted to other substances 350 13-2 Gluconeogenesis 354 Four gluconeogenic enzymes plus some glycolytic enzymes convert pyruvate to glucose 355 Gluconeogenesis is regulated at the fructose bisphosphatase step 356 CONTENTS 13-3 Glycogen Synthesis and Degradation 357 Glycogen synthesis consumes the free energy of UTP 358 Glycogen phosphorylase catalyzes glycogenolysis 359 13-4 The Pentose Phosphate Pathway 361 The oxidative reactions of the pentose phosphate pathway produce NADPH 361 Isomerization and interconversion reactions generate a variety of monosaccharides 362 A summary of glucose metabolism 363 BOX 13-A BIOCHEMISTRY NOTE Catabolism of Other Sugars 349 BOX 13-B CLINICAL CONNECTION Alcohol Metabolism 352 BOX 13-C CLINICAL CONNECTION Glycogen Storage Diseases 360 14 THE CITRIC ACID CYCLE 370 14-1 The Pyruvate Dehydrogenase Reaction 371 The pyruvate dehydrogenase complex contains multiple copies of three different enzymes 371 Pyruvate dehydrogenase converts pyruvate to acetyl-CoA 372 14-2 The Eight Reactions of the Citric Acid Cycle 374 1. Citrate synthase adds an acetyl group to oxaloacetate 375 2. Aconitase isomerizes citrate to isocitrate 376 3. Isocitrate dehydrogenase releases the first CO2 377 4. ␣-Ketoglutarate dehydrogenase releases the second CO2 378 5. Succinyl-CoA synthetase catalyzes substrate-level phosphorylation 378 6. Succinate dehydrogenase generates ubiquinol 379 7. Fumarase catalyzes a hydration reaction 380 8. Malate dehydrogenase regenerates oxaloacetate 380 The citric acid cycle is an energy-generating catalytic cycle 380 The citric acid cycle is regulated at three steps 381 The citric acid cycle probably evolved as a synthetic pathway 382 14-3 Anabolic and Catabolic Functions of the Citric Acid Cycle 384 Citric acid cycle intermediates are precursors of other molecules 384 Anaplerotic reactions replenish citric acid cycle intermediates 385 BOX 14-A CLINICAL CONNECTION Mutations in Citric Acid Cycle Enzymes 382 BOX 14-B BIOCHEMISTRY NOTE The Glyoxylate Pathway 386 15 OXIDATIVE PHOSPHORYLATION xi 394 15-1 The Thermodynamics of Oxidation–Reduction Reactions 395 Reduction potential indicates a substance's tendency to accept electrons 396 The free energy change can be calculated from the change in reduction potential 397 15-2 Mitochondrial Electron Transport 399 Mitochondrial membranes define two compartments 399 Complex I transfers electrons from NADH to ubiquinone 401 Other oxidation reactions contribute to the ubiquinol pool 403 Complex III transfers electrons from ubiquinol to cytochrome c 403 Complex IV oxidizes cytochrome c and reduces O2 406 15-3 Chemiosmosis 408 Chemiosmosis links electron transport and oxidative phosphorylation 408 The proton gradient is an electrochemical gradient 409 15-4 ATP Synthase 410 ATP synthase rotates as it translocates protons 410 The binding change mechanism explains how ATP is made 412 The P:O ratio describes the stoichiometry of oxidative phosphorylation 412 The rate of oxidative phosphorylation depends on the rate of fuel catabolism 413 BOX 15-A BIOCHEMISTRY NOTE Free Radicals and Aging 407 BOX 15-B BIOCHEMISTRY NOTE Uncoupling Agents Prevent ATP Synthesis 413 16 PHOTOSYNTHESIS 420 16-1 Chloroplasts and Solar Energy 422 Pigments absorb light of different wavelengths 422 Light-harvesting complexes transfer energy to the reaction center 424 16-2 The Light Reactions 425 Photosystem II is a light-activated oxidation–reduction enzyme 425 The oxygen-evolving complex of Photosystem II oxidizes water 426 Cytochrome b6 f links Photosystems I and II 428 A second photooxidation occurs at Photosystem I 428 Chemiosmosis provides the free energy for ATP synthesis 431 16-3 Carbon Fixation 432 Rubisco catalyzes CO2 fixation 432 The Calvin cycle rearranges sugar molecules 434 xii CONTENTS The availability of light regulates carbon fixation 436 Calvin cycle products are used to synthesize sucrose and starch 436 BOX 16-A BIOCHEMISTRY NOTE The C4 Pathway 433 17 LIPID METABOLISM 442 17-1 Fatty Acid Oxidation 445 Fatty acids are activated before they are degraded 445 Each round of ␤ oxidation has four reactions 446 Degradation of unsaturated fatty acids requires isomerization and reduction 449 Oxidation of odd-chain fatty acids yields propionyl-CoA 450 Some fatty acid oxidation occurs in peroxisomes 450 17-2 Fatty Acid Synthesis 453 Acetyl-CoA carboxylase catalyzes the first step of fatty acid synthesis 454 Fatty acid synthase catalyzes seven reactions 455 Other enzymes elongate and desaturate newly synthesized fatty acids 457 Fatty acid synthesis can be activated and inhibited 459 Acetyl-CoA can be converted to ketone bodies 461 17-3 Synthesis of Other Lipids 463 Triacylglycerols and phospholipids are built from acyl-CoA groups 463 Cholesterol synthesis begins with acetyl-CoA 466 Cholesterol can be used in several ways 467 A summary of lipid metabolism 469 BOX 17-A BIOCHEMISTRY NOTE Fats, Diet, and Heart Disease 458 BOX 17-B CLINICAL CONNECTION Inhibitors of Fatty Acid Synthesis 460 18 NITROGEN METABOLISM 475 18-1 Nitrogen Fixation and Assimilation 476 Nitrogenase converts N2 to NH3 476 Ammonia is assimilated by glutamine synthetase and glutamate synthase 477 Transamination moves amino groups between compounds 478 18-2 Amino Acid Biosynthesis 480 Several amino acids are easily synthesized from common metabolites 481 Amino acids with sulfur, branched chains, or aromatic groups are more difficult to synthesize 482 Amino acids are the precursors of some signaling molecules 486 18-3 Nucleotide Biosynthesis 488 Purine nucleotide synthesis yields IMP and then AMP and GMP 488 Pyrimidine nucleotide synthesis yields UTP and CTP 489 Ribonucleotide reductase converts ribonucleotides to deoxyribonucleotides 490 Thymidine nucleotides are produced by methylation 491 Nucleotide degradation produces uric acid or amino acids 492 18-4 Amino Acid Catabolism 494 Amino acids are glucogenic, ketogenic, or both 494 18-5 Nitrogen Disposal: The Urea Cycle 498 Glutamate supplies nitrogen to the urea cycle 499 The urea cycle consists of four reactions 499 BOX 18-A BIOCHEMISTRY NOTE Transaminases in the Clinic 480 BOX 18-B BIOCHEMISTRY NOTE Glyphosate, the Most Popular Herbicide 484 BOX 18-C BIOCHEMISTRY NOTE Nitric Oxide 487 BOX 18-D BIOCHEMISTRY NOTE Inborn Errors of Metabolism 497 19 REGULATION OF MAMMALIAN FUEL METABOLISM 509 19-1 Integration of Fuel Metabolism 510 Organs are specialized for different functions 510 Metabolites travel between organs 513 19-2 Hormonal Control of Fuel Metabolism 515 Insulin is released in response to glucose 515 Insulin promotes fuel use and storage 516 Glucagon and epinephrine trigger fuel mobilization 517 Additional hormones influence fuel metabolism 518 AMP-dependent protein kinase acts as a fuel sensor 519 19-3 Disorders of Fuel Metabolism 520 The body generates glucose and ketone bodies during starvation 520 Obesity has multiple causes 521 Diabetes is characterized by hyperglycemia 522 The metabolic syndrome links obesity and diabetes 523 BOX 19-A CLINICAL CONNECTION Cancer Metabolism 512 BOX 19-B BIOCHEMISTRY NOTE The Intestinal Microbiome Contributes to Metabolism 513 BOX 19-C BIOCHEMISTRY NOTE Marasmus and Kwashiorkor 520 CONTENTS part four 20 GENETIC INFORMATION DNA REPLICATION AND REPAIR 529 20-1 DNA Supercoiling 530 Topoisomerases alter DNA supercoiling 531 20-2 The DNA Replication Machinery 533 Replication occurs in factories 533 Helicases convert double-stranded DNA to singlestranded DNA 533 DNA polymerase faces two problems 534 DNA polymerases share a common structure and mechanism 536 DNA polymerase proofreads newly synthesized DNA 537 An RNase and a ligase are required to complete the lagging strand 538 20-3 Telomeres 540 Telomerase extends chromosomes 540 Is telomerase activity linked to cell immortality? 543 20-4 DNA Damage and Repair 543 DNA damage is unavoidable 543 Repair enzymes restore some types of damaged DNA 547 Base excision repair corrects the most frequent DNA lesions 548 Nucleotide excision repair targets the second most common form of DNA damage 549 Double-strand breaks can be repaired by joining the ends 550 Recombination also restores broken DNA molecules 551 20-5 DNA Packaging 552 The fundamental unit of DNA packaging is the nucleosome 552 Histones are covalently modified 553 DNA also undergoes covalent modification 554 BOX 20-A BIOCHEMISTRY NOTE HIV Reverse Transcriptase 541 BOX 20-B CLINICAL CONNECTION Cancer Is a Genetic Disease 544 21 xiii TRANSCRIPTION AND RNA 561 21-1 Transcription Initiation 563 Chromatin remodeling may precede transcription 563 Transcription begins at promoters 564 Transcription factors recognize eukaryotic promoters 566 Enhancers and silencers act at a distance from the promoter 566 Prokaryotic operons allow coordinated gene expression 570 21-2 RNA Polymerase 571 RNA polymerase is a processive enzyme 573 Transcription elongation requires a conformational change in RNA polymerase 573 Transcription is terminated in several ways 574 21-3 RNA Processing 576 Eukaryotic mRNAs receive a 5♣ cap and a 3♣ poly(A) tail 576 Splicing removes introns from eukaryotic genes 576 mRNA turnover and RNA interference limit gene expression 579 rRNA and tRNA processing includes the addition, deletion, and modification of nucleotides 581 BOX 21-A BIOCHEMISTRY NOTE DNA Binding Proteins 567 BOX 21-B BIOCHEMISTRY NOTE RNA: A Versatile Molecule 582 22 PROTEIN SYNTHESIS 590 22-1 tRNA Aminoacylation 592 tRNA aminoacylation consumes ATP 593 Some synthetases have proofreading activity 594 tRNA anticodons pair with mRNA codons 594 22-2 Ribosome Structure 596 22-3 Translation 599 Initiation requires an initiator tRNA 599 The appropriate tRNAs are delivered to the ribosome during elongation 601 The peptidyl transferase active site catalyzes peptide bond formation 603 Release factors mediate translation termination 606 Translation is efficient in vivo 606 22-4 Post-translational Events 608 Chaperones promote protein folding 608 The signal recognition particle targets some proteins for membrane translocation 609 Many proteins undergo covalent modification 611 BOX 22-A BIOCHEMISTRY NOTE The Genetic Code Expanded 595 BOX 22-B BIOCHEMISTRY NOTE Antibiotic Inhibitors of Protein Synthesis 604 GLOSSARY SOLUTIONS 617 629 ANSWERS TO PRACTICE PROBLEMS ANSWERS TO CLINICAL CONNECTION QUESTIONS CC-1 INDEX I-1 PP-1 APPLICATIONS AND ONLINE EXERCISES Many fascinating aspects of biochemistry do not fit easily into the pages of a textbook. For this reason, some material has been placed in boxes that do not interrupt the flow of the text.

State University Minnesota Tammy Stobb, St. Cloud State University Mississippi Jeffrey Evans, University of Southern Mississippi James R. Heitz, Mississippi State University Texas James Caras, Austin, Texas Paige Caras, Austin, Texas David W. Eldridge, Baylor University Edward Funkhouser, Texas A&M University Barrie Kitto, University of Texas at Austin Richard Sheardy, Texas Woman's University Linette Watkins, Southwest Texas State University Wisconsin Sandy Grunwald, University of Wisconsin La Crosse Canada Isabelle Barrette-Ng, University of Calgary Many of the molecular graphics that illustrate this book were created using publicly available coordinates from the Protein Data Bank (www.rcsb.org). The figures were rendered with the Swiss-Pdb Viewer [Guex, N., Peitsch, M.C., SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling, Electrophoresis 18, 2714–2723 (1997); program available at spdbv.vital-it.ch/] and Pov-Ray (available at www.povray.org) or with PyMol [developed by Warren DeLano; available at www.pymol.org/]. ACKNOWLEDGMENTS xxv GUIDE TO MEDIA RESOURCES Chapter Media Type Title Text Reference Chapter 1 The Chemical Basis of Life Bioinformatics Project 1. The Biochemical Literature page 19 Chapter 2 Aqueous Chemistry Animated Figure Titration curves for acetic acid, phosphate, and ammonia Figure 2.17 Chapter 3 From Genes to Proteins Interactive Exercise Guided Exploration Guided Exploration Three-dimensional structure of DNA Overview of transcription and translation DNA sequence determination by the chainterminator method PCR and site-directed mutagenesis Construction of a recombinant DNA molecule Site-directed mutagenesis 2. Databases for the Storage and "Mining" of Genome Sequences Figure 3-4 Section 3-2 Section 3-4 Guided Exploration Animated Figure Animated Figure Bioinformatics Project Chapter 4 Protein Structure Guided Exploration Animated Figure Guided Exploration Animated Figure Guided Exploration Animated Figure Animated Figure Guided Exploration Animated Figure Bioinformatics Project Chapter 5 Protein Function Animated Figure Animated Figure Animated Figure Animated Figure Animated Figure Bioinformatics Project Section 3-4 Figure 3-19 Figure 3-22 page 81 Stable helices in proteins: the ␣ helix The ␣ helix Hydrogen bonding in ␣ sheets ␣ sheets Secondary structrues in proteins Gel filtration chromatography Ion exchange chromatography Protein sequence determination Generating overlapping peptides to determine amino acid sequence 3. Visualizing Three-Dimensional Protein Structures Section 4-2 Figure 4-4 Section 4-2 Figure 4-6 Section 4-2 Figure 4-20 Figure 4-21 Section 4-5 Section 4-5 Oxygen-binding curve of hemoglobin Movements of heme and F helix in hemoglobin The Bohr effect Effect of BPG and CO2 on hemoglobin Mechanism of force generation in muscle 4. Using Databases to Compare and Identify Related Protein Seqences Figure 5-7 Figure 5-8 Figure 5-10 Figure 5-11 Figure 5-35 page 115 page 151 Chapter 6 How Enzymes Work Guided Exploration Animated Figure Catalytic mechanism of serine proteases Effect of preferential transition state binding Section 6-3 Figure 6-11 Chapter 7 Enzyme Kinetics and Inhibition Guided Exploration Section 7-3 Animated Figure Animated Figure Animated Figure Animated Figure Bioinformatics Project Michaelis–Menten kinetics, Lineweaver–Burk plots, and enzyme inhibition Progress curves for an enzyme-catalyzed reaction Plot of initial velocity versus substrate concentration Lineweaver–Burk plot Lineweaver–Burk plot of competitive inhibition 5. Enzyme Inhibitors and Rational Drug Design Figure 7-4 Figure 7-5 Figure 7-6 Figure 7-10 page 212 Chapter 8 Lipids and Membranes Guided Exploration Membrane structure and the fluid mosaic model Section 8-4 Chapter 9 Membrane Transport Interactive Exercise Animated Figure The K+ channel selectivity filter Model for glucose transport Figure 9-6 Figure 9-12 Chapter 10 Signaling Guided Exploration Interactive Exercise Interactive Exercise Hormone signaling by the adenylate cyclase system A heterotrimeric G protein C subunit of protein kinase A Section 10-2 Figure 10-4 Figure 10-6 xxvi GUIDE TO MEDIA RESOURCES xxvii Animated Figure Guided Exploration Interactive Exercise Interactive Exercise The phosphoinositide signaling system Hormone signaling by the receptor tyrosine kinase system Tyrosine kinase domain of the insulin receptor Glucocorticoid receptor DNA-binding domain in complex with DNA Section 10-2 Section 10-3 Figure 10-10 Figure 10-13 Chapter 12 Metabolism and Bioenergetics Bioinformatics Project 6. Metabolic Enzymes, Microarrays, and Proteomics page 332 Chapter 13 Glucose Metabolism Guided Exploration Animated Figure Animated Figure Interactive Exercise Animated Figure Section 13-1 Figure 13-2 Figure 13-4 Figure 13-5 Figure 13-6 Animated Figure Animated Figure Glycolysis overview Overview of glycolysis Mechanism of aldolase Triosephosphate isomerase Mechanism of glyceraldehyde-3-phosphate dehydrogenase Comparison of gluconeogenesis and glycolysis Overview of glucose metabolism Figure 13-10 Figure 13-12 Chapter 14 The Citric Acid Cycle Animated Figure Guided Exploration Interactive Exercise Overview of oxidative fuel metabolism Citric acid cycle overview Conformational changes in citrate synthase Figure 14-6 Section 14-2 Figure 14-7 Chapter 15 Oxidative Phosphorylation Guided Exploration Section 15-2 Interactive Exercise Interactive Exercise Guided Exploration Animated Figure Electron transport and oxidative phosphorylation overview Complex III Cytochrome c oxidase ATP synthase The binding change mechanism of ATP synthesis Figure 15-13 Figure 15-17 Section 15-4 Figure 15-25 Chapter 16 Photosynthesis Interactive Exercise Guided Exploration Animated Figure Light-harvesting complex LH-2 Two-center photosynthesis overview Mechanism of RUBP carboxylase Figure 16-7 Figure 16-24 Chapter 17 Lipid Metabolism Animated Figure Oxidation pathway Figure 17-6 Chapter 18 Nitrogen Metabolism Animated Figure Interactive Exercise Interactive Exercise Mechanism of PLP-dependent transamination The bifunctional enzyme tryptophan synthase E. coli ribonucleotide reductase Figure 18-5 Figure 18-8 Section 18-3 Chapter 19 Regulation of Mammalian Fuel Metabolism Animated Figure Animated Figure Animated Figure The Cori cycle The glucose–alanine cycle GLUT4 activity Figure 19-3 Figure 19-4 Figure 19-8 Chapter 20 DNA Replication and Repair Animated Figure Guided Exploration Guided Exploration Guided Exploration Interactive Exercise Interactive Exercise Interactive Exercise Guided Exploration Meselson and Stahl experiment DNA supercoiling DNA replication in E. coli E. coli DNA polymerase I PCNA HIV reverse transcriptase Human p53 in complex with its target DNA Nucleosome structure Figure 20-1 Section 20-1 Section 20-2 Figure 20-11 Figure 20-12 Box 20-A Box 20-B Section 20-5 Chapter 21 Transcription and RNA Interactive Exercise Guided Exploration Guided Exploration Figure 21-6 Box 21-A Section 21-1 Interactive Exercise Interactive Exercise TATA-binding protein in complex with TATA box Transcription factor–DNA interactions Regulation of gene expression by the lac repressor system An RNA–DNA hybrid helix The RNA polymerase II elongation complex Guided Exploration Interactive Exercise Guided Exploration Guided Exploration The structure of tRNA Ribosome with tRNA and mRNA Translational initiation Translational elongation Section 22-1 Figure 22-8 Section 22-3 Section 22-3 Chapter 22 Protein Synthesis Section 21-2 Figure 21-13 FOR THE STUDENT: HOW TO USE THIS BOOK Welcome to Biochemistry! Your success in this course will depend to a great extent on your willingness to take an active role in your education. Learning biochemistry requires more than simply reading the textbook, although we recommend that as a first step! Essential Biochemistry, Third Edition has been designed and written with you in mind, and we urge you to take advantage of all it has to offer. Biochemical knowledge is cumulative; it is not something that can be learned all at once. We advise you to keep up with your reading and other assignments so that you have plenty of time to reflect, ask questions, and, if necessary, seek help from your instructor. As you read each chapter of the textbook, make sure you understand how it fits into the course syllabus. Use the study aids provided in the textbook: First check out the Do You Remember questions to be sure you are prepared. Note the list of Key Concepts at the start of each text section. Use the Concept Review and Chapter Summary to check your understanding. Take time to answer the figure questions and the questions at the ends of Biochemistry Note and Clinical Connection boxes. Be sure to view the online resources that expand on material covered in the textbook. These exercises include animations of dynamic biochemical processes and interactive molecular graphics. You can enrich your understanding of biochemistry by exploring the exercises and answering the questions they pose. As you study, note the key sentences that are highlighted in italics. Be able to define the key terms in boldface, and test your knowledge by answering the questions in WileyPLUS. Most importantly, solve the problems at the end of each chapter. You should make every effort to complete all the problems without looking at the solutions. Developing problem-solving skills will facilitate your understanding of biochemistry and will help pave the way to success in any future academic or career endeavor. Finally, take advantage of the additional resources available—such as the list of Selected Readings, Bioinformatics Exercises, and articles from the Wiley Encyclopedia of Chemical Biology—if you need help, are curious about biochemistry, or need up-to-date information as a starting point for a class project. In writing Essential Biochemistry, we endeavored to select topics that would provide a solid introduction to modern biochemistry, which is a vast and ever-changing field. We realize that most students who use this book will not become biochemists. Nevertheless, it is our hope that you will come to understand the major themes of biochemistry and see how they relate to current and future developments in science and medicine. Charlotte W. Pratt Kathleen Cornely xxviii Third Edition Essential Biochemistry This page is intentionally left blank chapter 1 THE CHEMICAL BASIS OF LIFE DO living systems obey the laws of thermodynamics? The rules that describe matter and energy suggest that chemical systems tend to reach a state of equilibrium, in the process becoming more disordered. Yet we know that living systems, such as the Anabaena cells shown here, are not at equilibrium and are highly ordered. In this chapter, we'll introduce the molecular components of cells and explore the laws that apply to the chemical reactions that occur in cells, so you'll see that living systems obey the same rules that govern nonliving systems. [© Biophoto Associates/Photo Researchers, Inc.] THIS CHAPTER IN CONTEXT Part 1 Foundations 1 The Chemical Basis of Life Part 2 Molecular Structure and Function Part 3 Metabolism Part 4 Genetic Information 1 2 Ch 1 The Chemical Basis of Life This first chapter offers a preview of the study of biochemistry, broken down into three sections that reflect how topics in this book are organized. First come brief descriptions of the four major types of small biological molecules and their polymeric forms. Next is a summary of the thermodynamics that apply to metabolic reactions. Finally, there is a discussion of the origin of self-replicating life-forms and their evolution into modern cells. These short discussions introduce some of the key players and major themes of biochemistry and provide a foundation for the topics that will be encountered in subsequent chapters. 1-1 What Is Biochemistry? Biochemistry is the scientific discipline that seeks to explain life at the molecular level. It uses the tools and terminology of chemistry to describe the various attributes of living organisms. Biochemistry offers answers to such fundamental questions as "What are we made of ?" and "How do we work?" Biochemistry is also a practical science: It generates powerful techniques that underlie advances in other fields, such as genetics, cell biology, and immunology; it offers insights into the treatment of diseases such as cancer and diabetes; and it improves the efficiency of industries such as wastewater treatment, food production, and drug manufacturing. Biochemistry has traditionally been a science of reductionism; that is, it attempts to explain the whole by breaking it into smaller parts and examining each part separately. For biochemists, this means isolating and characterizing an organism's component molecules (Fig. 1-1). A thorough understanding of each molecule's physical structure and chemical reactivity helps lead to an understanding of how molecules cooperate and combine to form larger functional units and, ultimately, the intact organism. Nevertheless, a holistic approach is indispensable for unraveling the secrets of nature. Just as a clock completely disassembled no longer resembles a clock, information about a multitude of biological molecules does not necessarily reveal how an organism lives. Some molecular interactions are too complex to be teased apart in the laboratory, so it may be necessary to examine a cultured organism to see how it fares when a particular molecule is modified or absent. In addition, so much is known about so many molecules that the sheer volume of data—much of it collected in online databases—requires a computer to access and analyze it using the tools of Figure 1-1 Levels of organization Organism in a living organism. Biochemistry focuses on the structures and functions of molecules. Interactions between molecules give rise to higher-order structures (for example, organelles), which may themselves be components of larger entities, leading ultimately to the entire organism. Organ Cell Organelle Liver Hepatocyte Mitochondrion Molecules Human Citrate synthase DNA Citrate Ubiquinone bioinformatics (see Bioinformatics Project 1, The Biochemical Literature). Thus, a biochemist's laboratory is as likely to hold racks of test tubes as flasks of bacteria or computers. The material included in this book reflects both traditional and modern approaches to biochemical understanding. Chapters 3 through 22 of this book are divided into three groups that roughly correspond to three major themes of biochemistry: 3 Biological Molecules 1. What are living organisms made of ? Some molecules are responsible for the physical shapes of cells. Others carry out various activities in the cell. (For convenience, we use cell interchangeably with organism since the simplest living entity is a single cell.) In all cases, the structure of a molecule is intimately linked to its function. Understanding a molecule's structural characteristics is therefore an important key to understanding its functional significance. 2. How do organisms acquire and use energy? The ability of a cell to carry out metabolic reactions—to synthesize its constituents and to move, grow, and reproduce— requires the input of energy. A cell must extract this energy from the environment and spend it or store it in a manageable form. 3. How does an organism maintain its identity across generations? Modern human beings look much like they did 100,000 years ago. Certain bacteria have persisted for millions, if not billions, of years. In all organisms, information specifying a cell's structural composition and functional capacity must be faithfully maintained and transmitted each time the cell divides. Even within its own lifetime, a cell may dramatically alter its shape or metabolic activities, but it does so within certain limits. Throughout this book we will examine how regulatory processes define an organism's ability to respond to changing internal and external conditions. In addition, we will examine the diseases that can result from a molecular defect in any of the cell's attributes—its structural components, its metabolism, or its ability to follow genetic instructions. 1-2 Biological Molecules Even the simplest organisms contain a staggering number of different molecules, yet this number represents only an infinitesimal portion of all the molecules that are chemically possible. For one thing, only a small subset of the known elements are found in living systems (Fig. 1-2). The most abundant of these are C, N, O, and H, followed by Ca, P, K, S, Cl, Na, and Mg. Certain trace elements are also present in very small quantities. Virtually all the molecules in a living organism contain carbon, so biochemistry can be considered to be a branch of organic chemistry. In addition, biological molecules are constructed from H, N, O, P, and S. Most of these molecules belong to one of a few structural classes, which are described below. Similarly, the chemical reactivity of biomolecules is limited relative to the reactivity of all chemical compounds. A few of the functional groups and intramolecular linkages that are common in biochemistry are listed in Table 1-1. It may be helpful to review them, because we will refer to them throughout this book. Figure 1-2 Elements found in biological systems. The most abundant 1 H 11 Na 19 K KEY CONCEPTS • Biological molecules are composed of a subset of all possible elements and functional groups. • Cells contain four major types of biological molecules and three major types of polymers. 12 Mg 20 Ca 5 B 13 Al 23 V 24 25 26 27 28 29 30 Cr Mn Fe Co Ni Cu Zn 42 48 Mo Cd 74 W 6 C 14 Si 7 N 15 P 33 As 8 O 16 S 34 Se 9 F 17 Cl 35 Br 53 I elements are most darkly shaded; trace elements are most lightly shaded. Not every organism contains every trace element. Biological molecules primarily contain H, C, N, O, P, and S. TABLE 1-1 Common Functional Groups and Linkages in Biochemistry Compound Name Structurea Amineb RNH2 or R2NH or R3N or Functional Group RNH⩽ 3 R2NH⩽ 2 R3NH⩽ N ⩽ or N (amino group) Alcohol ROH OH (hydroxyl group) Thiol RSH SH (sulfhydryl group) Ether ROR O Aldehyde Ketone R R O O C H C O O C R (carbonyl group), R C O (carbonyl group), R C R C O OH or C O OH (carboxyl group) or R C O⩾ C O⩾ (carboxylate group) R C (acyl group) (acyl group) O O Ester O C O Carboxylic acidb (Carboxylate) (ether linkage) O OR C O (ester linkage) C N (amido group) C N or O Amide Imineb R C O NH2 R C O NHR R C NR2 R R NH or NR or O R R NH⩽ 2 NHR⩽ O Phosphoric acid esterb R O P O P OH or O O P O⩾ P O P O⩾ O OH or P O O P O OH or P P O O OH OH O O O O⩾ P P O O (phosphoanhydride linkage) OH O O O⩾ O⩾ (phosphoryl group, Pi) O⩾ OH OH O R (phosphoester linkage) O OH O R P O O⩾ Diphosphoric acid esterb H (imino group) O OH O R ⩽ C N P O OH or P O O P O⩾ O⩾ OH O⩾ OH (diphosphoryl group, pyrophosphoryl group, PPi) a R represents any carbon-containing group. In a molecule with more than one R group, the groups may be the same or different. b Under physiological conditions, these groups are ionized and hence bear a positive or negative charge. ? Cover the Structure column and draw the structure for each compound listed on the left. Do the same for each functional group. Cells contain four major types of biomolecules Most of the cell's small molecules can be divided into four classes. Although each class contains many members, they are united under a single structural or functional definition. Identifying a particular molecule's class may help predict its chemical properties and possibly its role in the cell. COO⩾ H C CH3 NH⩽ 3 (a) 1. Amino Acids The simplest compounds are the amino acids, so named because they contain an amino group (ONH2) and a carboxylic acid group (OCOOH). Under physiologi2 cal conditions, these groups are actually ionized to ONH1 3 and OCOO . The common amino acid alanine—like other small molecules—can be depicted in different ways, for example, by a structural formula, a ball-and-stick model, or a spacefilling model (Fig. 1-3). Other amino acids resemble alanine in basic structure, but instead of a methyl group (OCH3), they have another group—called a side chain or R group—that may also contain N, O, or S; for example, (b) ⩾ COO H C ⩾ COO O CH2 C H NH2 NH⩽ 3 C CH2 SH NH⩽ 3 Asparagine Cysteine 2. Carbohydrates Simple carbohydrates (also called monosaccharides or just sugars) have the formula (CH2O)n , where n is $ 3. Glucose, a monosaccharide with six carbon atoms, has the formula C6H12O6. It is sometimes convenient to draw it as a ladder-like chain (left); however, glucose forms a cyclic structure in solution (right): H O C H C OH HO C H H C OH H C OH H HO CH2OH CH2OH O H OH H H H OH OH Glucose In the representation of the cyclic structure, the darker bonds project in front of the page and the lighter bonds project behind it. In many monosaccharides, one or more hydroxyl groups are replaced by other groups, but the ring structure and multiple OOH groups of these molecules allow them to be easily recognized as carbohydrates. 3. Nucleotides A five-carbon sugar, a nitrogen-containing ring, and one or more phosphate groups are the components of nucleotides. For example, adenosine triphosphate (ATP) (c) Figure 1-3 Representations of alanine. The structural formula (a) indicates all the atoms and the major bonds. Some bonds, such as the COO and NOH bonds, are implied. Because the central carbon atom has tetrahedral geometry, its four bonds do not lie flat in the plane of the paper: The horizontal bonds actually extend slightly above the plane of the page, and the vertical bonds extend slightly behind it. This tetrahedral arrangement is more accurately depicted in the ball-and-stick model (b). Here, the atoms are color-coded by convention: C is gray, N is blue, O is red, and H is white. This ball-and-stick representation reveals the identities of the atoms and their positions in space but does not indicate their relative size or electrical charge. In a space-filling model (c), each atom is presented as a sphere whose radius corresponds to the distance of closest approach by another atom. This model most accurately depicts the actual size of the molecule, but it obscures some of its atoms and linkages. contains the nitrogenous group adenine linked to the monosaccharide ribose, to which a triphosphate group is also attached: 6 Ch 1 The Chemical Basis of Life NH2 Triphosphate O ⩾ O O P O O⩾ N N O P O O⩾ P O O⩾ H N N O CH2 Adenine H H OH OH Ribose H Adenosine triphosphate (ATP) The most common nucleotides are mono-, di-, and triphosphates containing the nitrogenous ring compounds (or "bases") adenine, cytosine, guanine, thymine, or uracil (abbreviated A, C, G, T, and U). 4. Lipids The fourth major group of biomolecules consists of the lipids. These compounds cannot be described by a single structural formula since they are a diverse collection of molecules. However, they all have in common a tendency to be poorly soluble in water because the bulk of their structure is hydrocarbon-like. For example, palmitic acid consists of a highly insoluble chain of 15 carbons attached to a carboxylic acid group, which is ionized under physiological conditions. The anionic lipid is therefore called palmitate. O CH2 CH2 H3C CH2 CH2 CH2 CH2 CH2 CH2 CH2 CH2 CH2 CH2 CH2 C O⩾ CH2 Palmitate Cholesterol, although it differs significantly in structure from palmitate, is also poorly soluble in water because of its hydrocarbon-like composition. CH3 CH CH2 CH3 CH2 CH2 CH3 CH CH3 CH3 HO Cholesterol Cells also contain a few other small molecules that cannot be easily classified into the groups above or that are constructed from molecules belonging to more than one group. There are three major kinds of biological polymers In addition to small molecules consisting of relatively few atoms, organisms contain macromolecules that may consist of thousands of atoms. Such huge molecules are not synthesized in one piece but are built from smaller units. This is a universal feature of BOX 1-A 7 Biological Molecules BIOCHEMISTRY NOTE Units Used in Biochemistry Convention rules the biochemist's choice of terms to quantify objects on a molecular scale. For example, the mass of a molecule can be expressed in atomic mass units; however, the masses of biological molecules—especially very large ones—are typically given without units. Here it is understood that the mass is expressed relative to one-twelfth the mass of an atom of the common carbon isotope 12C (12.011 atomic mass units). Occasionally, units of daltons (D) are used (1 dalton 5 1 atomic mass unit), often with the prefix kilo, k (kD). This is useful for macromolecules such as proteins, many of which have masses in the range from 20,000 (20 kD) to over 1,000,000 (1000 kD). The standard metric prefixes are also necessary for expressing the minute concentrations of biomolecules in living cells. Concentrations are usually given as moles per liter (mol ? L21 or M), with the appropriate prefix such as m, m, or n: mega (M) kilo (k) milli (m) micro (m) 106 103 1023 1026 nano (n) pico (p) femto (f) 1029 10212 10215 For example, the concentration of the sugar glucose in human blood is about 5 mM, but many intracellular molecules are present at concentrations of mM or less. Distances are customarily expressed in angstroms, Å (1 Å 5 10210 m) or in nanometers, nm (1 nm 5 1029 m). For example, the distance between the centers of carbon atoms in a COC bond is about 1.5 Å, and the diameter of a DNA molecule is about 20 Å. Question: The diameter of a typical spherical bacterial cell is about 1 mM. What is the cell's volume? nature: A few kinds of building blocks can be combined in different ways to produce a wide variety of larger structures. This is advantageous for a cell, which can get by with a limited array of raw materials. In addition, the very act of chemically linking individual units (monomers) into longer strings (polymers) is a way of encoding information (the sequence of the monomeric units) in a stable form. Biochemists use certain units of measure to describe both large and small molecules (Box 1-A). Amino acids, monosaccharides, and nucleotides each form polymeric structures with widely varying properties. In most cases, the individual monomers become covalently linked in head-to-tail fashion: Residue Monomers Polymer The linkage between monomeric units is characteristic of each type of polymer. The monomers are called residues after they have been incorporated into the polymer. Strictly speaking, lipids do not form polymers, although they do tend to aggregate to form larger structures. 1. Proteins Polymers of amino acids are called polypeptides or proteins. Twenty different amino acids serve as building blocks for proteins, which may contain many hundreds of amino acid residues. The amino acid residues are linked to each other by 8 Ch 1 The Chemical Basis of Life amide bonds called peptide bonds. A peptide bond (arrow) links the two residues in a dipeptide (the side chains of the amino acids are represented by R1 and R2). R1 O ⩽ H3N C R2 C N H O C C O⩾ H H Because the side chains of the 20 amino acids have different sizes, shapes, and chemical properties, the exact conformation (three-dimensional shape) of the polypeptide chain depends on its amino acid composition and sequence. For example, the small polypeptide endothelin, with 21 residues, assumes a compact shape in which the polymer bends and folds to accommodate the functional groups of its amino acid residues (Fig. 1-4). The 20 different amino acids can be combined in almost any order and in almost any proportion to produce myriad polypeptides, all of which have unique threedimensional shapes. This property makes proteins as a class the most structurally variable and therefore the most functionally versatile of all the biopolymers. Accordingly, proteins perform a wide variety of tasks in the cell, such as mediating chemical reactions and providing structural support. (a) 2. Nucleic Acids Polymers of nucleotides are termed polynucleotides or nucleic acids, better known as DNA and RNA. Unlike polypeptides, with 20 different amino acids available for polymerization, each nucleic acid is made from just four different nucleotides. For example, the residues in RNA contain the bases adenine, cytosine, guanine, and uracil, whereas the residues in DNA contain adenine, cytosine, guanine, and thymine. Polymerization involves the phosphate and sugar groups of the nucleotides, which become linked by phosphodiester bonds. O⩾ ⩾ O P O O CH2 O (b) H Base H H H Figure 1-4 Structure of human endothelin. The 21 amino acid residues of this polypeptide, shaded from blue to red, form a compact structure. In (a), each amino acid residue is represented by a sphere. The ball-and-stick model (b) shows all the atoms except hydrogen. [Structure (pdb 1EDN) determined by B. A. Wallace and R. W. Jones.] O Phosphodiester bond ⩾O P H O O CH2 O H Base H H H OH H In part because nucleotides are much less variable in structure and chemistry than amino acids, nucleic acids tend to have more regular structures than proteins. This is in keeping with their primary role as carriers of genetic information, which is contained in their sequence of nucleotide residues rather than in their three-dimensional shape (Fig. 1-5). Nevertheless, some nucleic acids do bend and fold into compact globular shapes, as proteins do. 3. Polysaccharides Polysaccharides usually contain only one or a few different types of monosaccharide residues, so even though a cell may synthesize dozens of different kinds of monosaccharides, most of its polysaccharides are homogeneous polymers. This tends to limit their potential for carrying genetic information in the sequence of their residues (as nucleic acids do) or for adopting a large variety of shapes and mediating chemical reactions (as proteins do). On the other hand, polysaccharides perform essential cell functions by serving as fuel-storage molecules and by providing structural support. For example, plants link the monosaccharide glucose, which is a fuel for virtually all cells, into the polysaccharide starch for long-term storage. The glucose residues are linked by glycosidic bonds (the bond is shown in red in this disaccharide): H HO CH2OH O H OH H H OH H H O CH2OH O H OH H H 9 Biological Molecules CGUACG (a) H OH OH Glucose monomers are also the building blocks for cellulose, the extended polymer that helps make plant cell walls rigid (Fig. 1-6). The starch and cellulose polymers differ in the arrangement of the glycosidic bonds between glucose residues. (b) The brief descriptions of biological polymers given above are generalizations, meant to convey some appreciation for the possible structures and functions of these macromolecules. Exceptions to the generalizations abound. For example, some small polysaccharides encode information that allows cells bearing the molecules on their Figure 1-5 Structure of a nucleic acid. (a) Sequence of nucleotide residues, using one-letter abbreviations. (b) Balland-stick model of the polynucleotide, showing all atoms except hydrogen (this structure is a six-residue segment of RNA). [Structure (pdb ARF0108) determined by R. Biswas, S. N. Mitra, and M. Sundaralingam.] Glucose Starch Cellulose Figure 1-6 Glucose and its polymers. Both starch and cellulose are polysaccharides containing glucose residues. They differ in the type of chemical linkage between the monosaccharide units. Starch molecules have a loose helical conformation, whereas cellulose molecules are extended and relatively stiff. 10 Ch 1 The Chemical Basis of Life TABLE 1-2 Biopolymer Functions of Biopolymers Encode Information Carry Out Metabolic Reactions — ✔ ✓ ✓ — Proteins Nucleic acids Polysaccharides ✔ Store Energy ✓ — ✔ Support Cellular Structures ✔ ✓ ✔ ✔ major function ✓ minor function surfaces to recognize each other. Likewise, some nucleic acids perform structural roles, for example, by serving as scaffolding in ribosomes, the small particles where protein synthesis takes place. Under certain conditions, proteins are called on as fuel-storage molecules. A summary of the major and minor functions of proteins, polysaccharides, and nucleic acids is presented in Table 1-2. CONCEPT REVIEW • Which six elements are most abundant in biological molecules? • Name the common functional groups and linkages shown in Table 1-1. • Give the structural or functional definitions for amino acids, monosaccharides, nucleotides, and lipids. • Why is it efficient for macromolecules to be polymers? • What is the relationship between a monomer and a residue? • Give the structural definitions and major functions of proteins, polysaccharides, and nucleic acids. 1-3 Energy and Metabolism KEY CONCEPTS • The free energy of a system is determined by its enthalpy and entropy. • Living organisms obey the laws of thermodynamics. Assembling small molecules into polymeric macromolecules requires energy. And unless the monomeric units are readily available, a cell must synthesize the monomers, which also requires energy. In fact, cells require energy for all the functions of living, growing, and reproducing. It is useful to describe the energy in biological systems using the terminology of thermodynamics (the study of heat and power). An organism, like any chemical system, is subject to the laws of thermodynamics. According to the first law of thermodynamics, energy cannot be created or destroyed. However, it can be transformed. For example, the energy of a river flowing over a dam can be harnessed as electricity, which can then be used to produce heat or perform mechanical work. Cells can be considered to be very small machines that use chemical energy to drive metabolic reactions, which may also produce heat or carry out mechanical work. Enthalpy and entropy are components of free energy The energy relevant to biochemical systems is called the Gibbs free energy (after the scientist who defined it) or just free energy. It is abbreviated G and has units of joules per mol (J ℷ mol⩾1). Free energy has two components: enthalpy and entropy. Enthalpy (abbreviated H, with units of J ℷ mol⩾1) is taken to be equivalent to the heat content of the system. Entropy (abbreviated S, with units of J ℷ K⩾1 ℷ mol⩾1) is a measure of how the energy is dispersed within that system. Entropy can therefore be considered to be a measure of the system's disorder or randomness, because the more ways a system's components can be arranged, the more dispersed its energy. For example, consider a pool table at the start of a game when all 15 balls are arranged in one neat triangle (a state of high order or low entropy). After play has begun, the balls are scattered across the table, which is now in a state of disorder and high entropy (Fig. 1-7). Free energy, enthalpy, and entropy are related by the equation G 5 H 2 TS [1-1] where T represents temperature in Kelvin (equivalent to degrees Celsius plus 273). Temperature is a coefficient of the entropy term because entropy varies with temperature; the entropy of a substance increases when it is warmed because more thermal energy has been dispersed within it. The enthalpy of a chemical system can be measured, although with some difficulty, but it is next to impossible to measure a system's entropy because this would require counting all the possible arrangements of its components or all the ways its energy could be spread out among them. Therefore, it is more practical to deal with changes in these quantities (change is indicated by the Greek letter delta, ⌬) so that DG 5 DH 2 T DS (a) (b) Figure 1-7 Illustration of entropy. [1-2] Biochemists can measure how the free energy, enthalpy, and entropy of a system differ before and after a chemical reaction. For example, some chemical reactions are accompanied by the release of heat to the surroundings (Hafter 2 Hbefore 5 DH , 0), whereas others absorb heat from the surroundings (DH . 0). Similarly, the entropy change, Safter 2 S before 5 DS, can be positive or negative. When ⌬H and ⌬S for a process are known, Equation 1-2 can be used to calculate the value of ⌬G at a given temperature (see Sample Calculation 1-1). Entropy is a measure of the dispersal of energy in a system, so it reflects the system's randomness or disorder. (a) Entropy is low when all the balls are arranged in a single area of the pool table. (b) Entropy is high after the balls have been scattered, because there are now a large number of different possible arrangements of the balls on the table. ? Compare the entropy of a ball of yarn before and after a cat has played with it. SAMPLE CALCULATION 1-1 Use the information below to calculate the change in enthalpy and the change in entropy for the reaction A S B. A B DH 5 5 5 5 Enthalpy (kJ ℷ mol⩾1) 60 75 HB 2 HA 75 kJ ? mol21 2 60 kJ ? mol21 15 kJ ? mol21 15,000 J ? mol21 PROBLEM Entropy (J ℷ K⩾1 ℷ mol⩾1) 22 97 DS 5 SB 2 SA 5 97 J ? K21 ? mol21 222 J ? K21 ? mol21 5 75 J ? K21 ? mol21 SOLUTION 1. The change in enthalpy for a reaction is 8 kJ ℷ mol⩾1. Is heat absorbed or given off during the reaction? 2. A reacting system undergoes a change from ordered to disordered. Is the entropy change positive or negative? ⌬G is less than zero for a spontaneous process A china cup dropped from a great height will break, but the pieces will never reassemble themselves to restore the cup. The thermodynamic explanation is that the broken pieces have less free energy than the intact cup. In order for a process to occur, the overall change in free energy (⌬G) must be negative. For a chemical reaction, this means that the free energy of the products must be less than the free energy of the reactants: DG 5 Gproducts 2 Greactants , 0 [1-3] PRACTICE PROBLEMS 12 Ch 1 The Chemical Basis of Life When ⌬G is less than zero, the reaction is said to be spontaneous or exergonic. A nonspontaneous or endergonic reaction has a free energy change greater than zero; in this case, the reverse reaction is spontaneous. ASB DG , 0 Nonspontaneous BSA DG , 0 Spontaneous Note that thermodynamic spontaneity does not indicate how fast a reaction occurs, only whether it will occur as written. (The rate of a reaction depends on other factors, such as the concentrations of the reacting molecules, the temperature, and the presence of a catalyst.) When a reaction, such as A S B, is at equilibrium, the rate of the forward reaction is equal to the rate of the reverse reaction, so there is no net change in the system. In this situation, ⌬G 0. A quick examination of Equation 1-2 reveals that a reaction that occurs with a decrease in enthalpy and an increase in entropy is spontaneous at all temperatures because ⌬G is always less than zero. These results are consistent with everyday experience. For example, heat moves spontaneously from a hot object to a cool object, and items that are neatly arranged tend to become disordered, never the other way around. (This is a manifestation of the second law of thermodynamics, which states that energy tends to spread out.) Accordingly, reactions in which the enthalpy increases and entropy decreases do not occur. If enthalpy and entropy both increase or both decrease during a reaction, the value of ⌬G then depends on the temperature, which governs whether the T ⌬S term of Equation 1-2 is greater than or less than the ⌬H term. This means that a large increase in entropy can offset an unfavorable (positive) change in enthalpy. Conversely, the release of a large amount of heat (⌬H ⬍ 0) during a reaction can offset an unfavorable decrease in entropy (see Sample Calculation 1-2). SAMPLE CALCULATION 1-2 PROBLEM Use the information given in Sample Calculation 1-1 to determine whether the reaction A S B is spontaneous at 25⬚C. SOLUTION Substitute the values for ⌬H and ⌬S, calculated in Sample Calculation 1-1, into Equation 1-2. To express the temperature in Kelvin, add 273 to the temperature in degrees Celsius: 273 ⩽ 25 298 K. DG 5 5 5 5 5 DH 2 T DS 15,000 J ? mol21 2 298 K (75 J ? K21 ? mol21) 15,000 2 22,400 J ? mol21 27400 J ? mol21 27.4 kJ ? mol21 Because ⌬G is less than zero, the reaction is spontaneous. Even though the change in enthalpy is unfavorable, the large increase in entropy makes ⌬G favorable. PRACTICE PROBLEMS 3. If ⌬H ⩾15,000 J ? mol⩾1 and ⌬S ⩾75 J ? K⩾1 ? mol⩾1, would the reaction be spontaneous at 25⬚C? 4. Use the information given in Sample Calculation 1-1 to determine the temperature at which ⌬G 0. Life is thermodynamically possible In order to exist, life must be thermodynamically spontaneous. Does this hold at the molecular level? When analyzed in a test tube (in vitro, literally "in glass"), many of a cell's metabolic reactions have free energy changes that are less than zero, but some reactions do not. Nevertheless, the nonspontaneous reactions are able to proceed in vivo (in a living organism) because they occur in concert with other reactions that B 13 Energy and Metabolism B Free energy (G) A C Reaction coordinate TABLE 1-3 Figure 1-8 Free energy changes in coupled reactions. A nonspontaneous reaction, such as A n B, which has a positive value of ⌬G, can be coupled to another reaction, B n C, which has a negative value of ⌬G and is therefore spontaneous. The reactions are coupled because the product of the first reaction, B, is a reactant for the second reaction. ? Which reaction occurs spontaneously in reverse: C n B, B n A, or C n A? are thermodynamically favorable. Consider two reactions in vitro, one nonspontaneous (⌬G ⬎ 0) and one spontaneous (⌬G ⬍ 0): 21 ASB BSC DG 5 115 kJ ? mol DG 5 220 kJ ? mol21 (nonspontaneous) (spontaneous) Compounda Formula Carbon dioxide most oxidized (least reduced) O Acetic acid H C O H C O C OH H When the reactions are combined, their ⌬G values are added, so the overall process has a negative change in free energy: A1BSB1C ASC Oxidation States of Carbon Carbon monoxide C O Formic acid H C O OH DG 5 (15 kJ ? mol21) 1 (220 kJ ? mol21) DG 5 25 kJ ? mol21 This phenomenon is shown graphically in Figure 1-8. In effect, the unfavorable "uphill" reaction A S B is pulled along by the more favorable "downhill" reaction B S C. Cells couple unfavorable metabolic processes with favorable ones so that the net change in free energy is negative. Note that it is permissible to add ⌬G values because the free energy, G, depends only on the initial and final states of the system, without regard to the specific chemical or mechanical work that occurred in going from one state to the other. Most macroscopic life on earth today is sustained by the energy of the sun (this was not always the case, nor is it true of all organisms). In photosynthetic organisms, such as green plants, light energy excites certain molecules so that their subsequent chemical reactions occur with a net negative change in free energy. These thermodynamically favorable (spontaneous) reactions are coupled to the unfavorable synthesis of monosaccharides from atmospheric CO2 (Fig. 1-9). In this process, the carbon is reduced. Reduction, the gain of electrons, is accomplished by the addition of hydrogen or the removal of oxygen (the oxidation states of carbon are reviewed in Table 1-3). The plant—or an animal that eats Acetone H H C H O C H Acetaldehyde H C O C H H H Formaldehyde C O H Acetylene H C C Ethanol H C C O C O Reduction (unfavorable) OH H H Ethene Free energy H H H H H Light energy H C H H C C H H H H H C OH Carbon of a monosaccharide Oxidation (favorable) O C O Figure 1-9 Reduction and reoxidation of carbon compounds. The sun provides the free energy to convert CO2 to reduced compounds such as monosaccharides. The reoxidation of these compounds to CO2 is thermodynamically spontaneous, so free energy can be made available for other metabolic processes. Note that free energy is not actually a substance that is physically released from a molecule. Ethane H C C H H H H Methane least oxidized (most reduced) a H C H H Compounds are listed in order of decreasing oxidation state of the red carbon atom. 14 Ch 1 The Chemical Basis of Life DO living systems obey the laws of thermodynamics? the plant—can then break down the monosaccharide to use it as a fuel to power other metabolic activities. In the process, the carbon is oxidized—it loses electrons through the addition of oxygen or the removal of hydrogen—and ultimately becomes CO2. The oxidation of carbon is thermodynamically favorable, so it can be coupled to energy-requiring processes such as the synthesis of building blocks and their polymerization to form macromolecules. Virtually all metabolic processes occur with the aid of catalysts called enzymes, most of which are proteins (a catalyst greatly increases the rate of a reaction without itself undergoing any net change). For example, specific enzymes catalyze the formation of peptide, phosphodiester, and glycosidic linkages during polymer synthesis. Other enzymes catalyze cleavage of these bonds to break the polymers into their monomeric units. A living organism—with its high level of organization of atoms, molecules, and larger structures—represents a state of low entropy relative to its surroundings. Yet the organism can maintain this thermodynamically unfavorable state as long as it continually obtains free energy from its food. This is the answer to the question posed at the start of the chapter: Living organisms do indeed obey the laws of thermodynamics. When the organism ceases to obtain a source of free energy from its surroundings or exhausts its stored food, the chemical reactions in its cells reach equilibrium (⌬G 0), which results in death. CONCEPT REVIEW • What do enthalpy and entropy mean and how are they related to free energy? • What does the value of ⌬G reveal about a biochemical process? • How can thermodynamically unfavorable reactions proceed in vivo? • Why must an organism have a steady supply of food? • Describe the cycle of carbon reduction and oxidation in photosynthesis and in the breakdown of a compound such as a monosaccharide. 1-4 The Origin and Evolution of Life KEY CONCEPTS • Modern prokaryotic and eukaryotic cells apparently evolved from simpler nonliving systems. • The three domains of life are bacteria, archaea, and eukarya. Every living cell originates from the division of a parental cell. Thus, the ability to replicate (make a replica or copy of itself) is one of the universal characteristics of living organisms. In order to leave descendants that closely resemble itself, a cell must contain a set of instructions—and the means for carrying them out—that can be transmitted from generation to generation. Over time, the instructions change gradually, so that species also change, or evolve. By carefully examining an organism's genetic information and the cellular machinery that supports it, biochemists can draw some conclusions about the organism's relationship to more ancient life-forms. The history of evolution is therefore contained not just within the fossil record but also in the molecular makeup of all living cells. For example, nucleic acids participate in the storage and transmission of genetic information in all organisms, and the oxidation of glucose is an almost universal means for generating metabolic free energy. Consequently, DNA, RNA, and glucose must have been present in the ancestor of all cells. The prebiotic world A combination of theory and experimental data leads to several plausible scenarios for the emergence of life from nonbiological (prebiotic) materials on the early earth. In one scenario, inorganic compounds such as H2, H2O, NH3, and CH4—which may have been present in the early atmosphere—could have given rise to simple biomolecules, such as amino acids, when sparked by lightning. Laboratory experiments with the same raw materials and electrical discharges to simulate lightning do 15 The Origin and Evolution of Life Electrodes Mixture of gases Condenser Figure 1-10 Laboratory synthesis of biological molecules. A mixture of Water gasesOH2, H2O, NH3, and CH4Ois subject to an electrical discharge. Newly formed compounds, such as amino acids, accumulate in the aqueous phase as water vapor condenses. Samples of the reaction products can be removed via the stopcock. Heat source Stopcock for removing samples in fact yield these molecules (Fig. 1-10). Other experiments suggest that hydrogen cyanide (HCN), formaldehyde (HCOH), and phosphate could have been converted to nucleotides with a similarly modest input of energy. Over time, simple molecular building blocks could have accumulated and formed larger structures, particularly in shallow waters where evaporation would have had a concentrating effect. Eventually, conditions would have been ripe for the assembly of functional, living cells. Charles Darwin proposed that life might have arisen in some "warm little pond"; however, the early earth was probably a much more violent place, with frequent meteorite impacts and volcanic activity. In an alternative scenario, supported by studies of the metabolism of some modern bacteria, the first cells could have developed at deep-sea hydrothermal vents, some of which are characterized by temperatures as high as 350⬚C and clouds of gaseous H2S and metal sulfides (giving them the name "black smokers"; Fig. 1-11). In the laboratory, incubating a few small molecules in the presence of iron sulfide and nickel sulfide at 100⬚C yields acetic acid, an organic compound with a newly formed COC bond: CH3SH Methyl thiol 1 CO FeS, NiS 1 H2O ¡ CH3COOH 1 H2S Carbon monoxide Acetic acid Under similar conditions, amino acids spontaneously form short polypeptides. Although the high temperatures that are necessary for their synthesis also tend to break them down, these compounds would have been stable in the cooler water next to the hydrothermal vent. Regardless of how they formed, the first biological building blocks would have had to polymerize. This process might have been stimulated when the organic molecules—often bearing anionic (negatively charged) groups—aligned themselves on a cationic (positively charged) mineral surface. Figure 1-11 A hydrothermal vent. Life − Monomers − + Polymer − − + + Clay + − − − − + + + + − − − − + + + + may have originated at these "black smokers," where high temperatures, H2S, and metal sulfides might have stimulated the formation of biological molecules. [B. Murton/Southhampton Oceanography Centre/Science Photo Library/ Photo Researchers.] 16 Ch 1 The Chemical Basis of Life A A A A A Original U U U U U A A A A A U U U U U In fact, in the laboratory, common clay promotes the polymerization of nucleotides into RNA. Primitive polymers would have had to gain the capacity for selfreplication. Otherwise, no matter how stable or chemically versatile, such molecules would never have given rise to anything larger or more complicated: The probability of assembling a fully functional cell from a solution of thousands of separate small molecules is practically nil. Because RNA in modern cells represents a form of genetic information and participates in all aspects of expressing that information, it may be similar to the first self-replicating biopolymer. It might have made a copy of itself by first making a complement, a sort of mirror image, that could then make a complement of itself, which would be identical to the original molecule (Fig. 1-12). 1. The polyA molecule serves as a template for the synthesis of a polymer containing uracil nucleotides, U, which are complementary to adenine nucleotides (in modern RNA, A pairs with U). 2. The two polymer chains separate. A A A A A U U U U U + Original Complement A A A A A A A A A A + U U U U U A A A A A U U U U U A A A A A + 3. The polyU molecule serves as a template for the synthesis of a new complementary polyA chain. 4. The chains again separate and the polyU polymer is discarded, leaving the original polyA molecule and its exact copy. A A A A A Original Copy Figure 1-12 Possible mechanism for the selfreplication of a primitive RNA molecule. For simplicity, the RNA molecule is shown as a polymer of adenine nucleotides, A. ? Draw a diagram showing how polyU would be replicated. Origins of modern cells A replicating molecule's chances of increasing in number depend on natural selection, the phenomenon whereby the entities best suited to the prevailing conditions are the likeliest to survive and multiply (Box 1-B). This would have favored a replicator that was chemically stable and had a ready supply of building blocks and free energy for making copies of itself. Accordingly, it would have been advantageous to become enclosed in some sort of membrane that could prevent valuable small molecules from diffusing away. Natural selection would also have favored replicating systems that developed the means for synthesizing their own building blocks and for more efficiently harnessing sources of free energy. The first cells were probably able to "fix" CO2—that is, convert it to reduced organic compounds—using the free energy released in the oxidation of readily available inorganic compounds such as H2S or Fe2⩽. Vestiges of these processes can be seen in modern metabolic reactions that involve sulfur and iron. Later, photosynthetic organisms similar to present-day cyanobacteria (also called blue-green algae) used the sun's energy to fix CO2: CO2 1 H2O n (CH2O) 1 O2 The concomitant oxidation of H2O to O2 dramatically increased the concentration of atmospheric O2, about 2.4 billion years ago, and made it possible for aerobic (oxygen-using) organisms to avail themselves of this powerful oxidizing agent. The anaerobic origins of life are still visible in the most basic metabolic reactions of modern organisms; these reactions proceed in the absence of oxygen. Now that the earth's atmosphere contains about 18% oxygen, anaerobic organisms have not disappeared, but they have been restricted to microenvironments where O2 is scarce, such as the digestive systems of animals or underwater sediments. The earth's present-day life-forms are of two types, which are distinguished by their cellular architecture: 1. Prokaryotes are small unicellular organisms that lack a discrete nucleus and usually contain no internal membrane systems. This group comprises two subgroups that are remarkably different metabolically, although they are similar in appearance: the eubacteria (usually just called bacteria), exemplified by E. coli, and the archaea (or archaebacteria), best known as organisms that inhabit extreme environments, although they are actually found almost everywhere (Fig. 1-13). BOX 1-B BIOCHEMISTRY NOTE 17 The Origin and Evolution of Life How Does Evolution Work? Documenting evolutionary change is relatively straightforward, but the mechanisms whereby evolution occurs are prone to misunderstanding. Populations change over time, and new species arise as a result of natural selection. Selection operates on individuals, but its effects can be seen in a population only over a period of time. Most populations are collections of individuals that share an overall genetic makeup but also exhibit small variations due to random alterations (mutations) in their genetic material as it is passed from parent to offspring. In general, the survival of an individual depends on how well suited it is to the particular conditions under which it lives. Individuals whose genetic makeup grants them the greatest rate of survival have more opportunities to leave offspring with the same genetic makeup. Consequently, their characteristics become widespread in a population, and, over time, the population appears to adapt to its environment. A species that is well suited to its environment tends to persist; a poorly adapted species fails to reproduce and therefore dies out. Because evolution is the result of random variations and changing probabilities for successful reproduction, it is inherently random and unpredictable. Furthermore, natural selection acts on the raw materials at hand. It cannot create something out of nothing but must operate in increments. For example, the insect wing did not suddenly appear in the offspring of a wingless parent but most likely developed bit by bit, over many generations, by modification of a gill or heat-exchange appendage. Each step of the wing's development would have been subject to natural selection, eventually making an individual that bore the appendage more likely to survive, perhaps by being able to first glide and then actually fly in pursuit of food or to evade predators. Although we tend to think of evolution as an imperceptibly slow process, occurring on a geological time scale, it is ongoing and accessible to observation and quantification in the laboratory. For example, under optimal conditions, the bacterium Escherichia coli requires only about 20 minutes to produce a new generation. In the laboratory, a culture of E. coli cells can progress through about 2500 generations in a year (in contrast, 2500 human generations would require about 60,000 years). Hence, it is possible to subject a population of cultured bacterial cells to some "artificial" selection—for example, by making an essential nutrient scarce—and observe how the genetic composition of the population changes over time as it adapts to the new conditions. Question: Why can't acquired (rather than genetic) characteristics serve as the raw material for evolution? 2. Eukaryotic cells are usually larger than prokaryotic cells and contain a nucleus and other membrane-bounded cellular compartments (such as mitochondria, chloroplasts, and endoplasmic reticulum). Eukaryotes may be unicellular or multicellular. This group (also called the eukarya) includes microscopic organisms as well as familiar macroscopic plants and animals (Fig. 1-14). By analyzing the sequences of nucleotides in certain genes that are present in all species, it is possible to construct a diagram that indicates how the bacteria, archaea, and eukarya are related. The number of sequence differences between two groups of organisms indicates how long ago they diverged from a common ancestor: Species with similar sequences have a longer shared evolutionary history than species with dissimilar sequences. This sort of analysis has produced the evolutionary tree shown in Figure 1-15. The evolutionary history of eukaryotes is complicated by the fact that eukaryotic cells exhibit characteristics of both bacteria and archaea. Eukaryotic cells also contain organelles that are almost certainly the descendants of free-living cells. Specifically, the chloroplasts of plant cells, which carry out photosynthesis, closely resemble the Figure 1-13 Prokaryotic cells. These single-celled Escherichia coli bacteria lack a nucleus and internal membrane systems. [E. Gray/Science Photo Library/Photo Researchers.] 18 Ch 1 The Chemical Basis of Life Prokaryotes Eukaryotes Animals Eukarya Fungi Archaea Bacteria Plants Figure 1-15 Evolutionary tree based on nucleotide sequences. This diagram reveals that the bacteria separated before the archaea and eukarya diverged. Note that the closely spaced fungi, plants, and animals are actually more similar to each other than are many groups of prokaryotes. [After Wheelis, M. L., Kandler, O., and Woese, C. R., Proc. Natl. Figure 1-14 A eukaryotic cell. The paramecium, a one-celled organism, contains a nucleus and other membranebounded compartments. [Dr. David Patterson/Science Photo Library/Photo Researchers.] ? Compare this cell to the bacterial cells shown in Figure 1-13. Acad. Sci. USA 89, 2930–2934 (1992).] photosynthetic cyanobacteria. The mitochondria of plant and animal cells, which are the site of much of the eukaryotic cell's aerobic metabolism, resemble certain bacteria. In fact, both chloroplasts and mitochondria contain their own genetic material and protein-synthesizing machinery. It is likely that an early eukaryotic cell developed gradually from a mixed population of prokaryotic cells. Over many generations of living in close proximity and sharing each other's metabolic products, some of these cells became incorporated within a single larger cell. This arrangement would account for the mosaic-like character of modern eukaryotic cells (Fig. 1-16). At some point, cells in dense populations might have traded their individual existence for a colonial lifestyle. This would have allowed for a division of labor as cells became specialized and would have eventually produced multicellular organisms. The earth currently sustains about 9 million different species (although estimates vary widely). Perhaps some 500 million species have appeared and vanished over the course of evolutionary history. It is unlikely that the earth harbors more than a few mammals that have yet to be discovered, but new microbial species are routinely described. And although the number of known prokaryotes (about 10,000) is much less than the number of known eukaryotes (for example, there are about 900,000 known species of insects), prokaryotic metabolic strategies are amazingly varied. Nevertheless, by documenting characteristics that are common to all species, we can derive far-reaching conclusions about what life is made of, what sustains it, and how it has developed over the eons. Nucleus DNA Organelles Figure 1-16 Possible origin of eukaryotic cells. The close association of different kinds of free-living cells gradually led to the modern eukaryotic cell, which appears to be a mosaic of bacterial and archaeal features and contains organelles that resemble whole bacterial cells. CONCEPT REVIEW • Explain how simple prebiotic compounds could give rise to biological monomers and polymers. • Why is replication a requirement for life? • Why were the first organisms anaerobic, and why are they now relatively scarce? • Describe the differences between prokaryotes and eukaryotes. • Why do eukaryotic cells appear to be mosaics? 19 SUMMARY 1-2 Biological Molecules • The most abundant elements in biological molecules are H, C, N, O, P, and S, but a variety of other elements are also present in living systems. • The major classes of small molecules in cells are amino acids, monosaccharides, nucleotides, and lipids. The major types of biological polymers are proteins, nucleic acids, and polysaccharides. 1-3 Energy and Metabolism • Life is thermodynamically possible because unfavorable endergonic processes are coupled to favorable exergonic processes. 1-4 The Origin and Evolution of Life • The earliest cells may have evolved in concentrated solutions of molecules or near hydrothermal vents. • Eukaryotic cells contain membrane-bounded organelles. Prokaryotic cells, which are smaller and simpler, include the bacteria and the archaea. • Free energy has two components: enthalpy (heat content) and entropy (disorder). Free energy decreases in a spontaneous process. GLOSSARY TERMS bioinformatics trace element amino acid carbohydrate monosaccharide nucleotide lipid monomer polymer residue polypeptide protein peptide bond conformation polynucleotide nucleic acid phosphodiester bond polysaccharide glycosidic bond free energy (G) enthalpy (H) entropy (S) ⌬G spontaneous process exergonic reaction nonspontaneous process endergonic reaction in vitro BIOINFORMATICS in vivo reduction oxidation enzyme replication complement natural selection aerobic anaerobic prokaryote bacteria archaea eukaryote eukarya Learn to use the PubMed database to search for published papers on a given topic and to retrieve information from published reports. PROJECT 1 THE BIOCHEMICAL LITERATURE PROBLEMS 1-2 Biological Molecules (c) O 1. Identify the functional groups in the following molecules: (a) O ⩽H 3N CH C (b) O⩾ C O O NH2 HO O O CH CH2 CH3 H C C (CH2)14 O O CH2 CH2 C CH2 COOH P OH OH OH C NH C H HO C H H C OH H C OH CH2OH O (d) ⩽H 3N CH C CH2 SH O⩾ CH3 20 (e) O H O (f) C H C OH O⩾ C HN CH2OH 2. The structures of several vitamins are shown below. Identify the functional groups in each vitamin. O OH O O C HO OH 5. The nutritive quality of food can be analyzed by measuring the chemical elements it contains. Most foods are mixtures of the three major types of molecules: fats (lipids), carbohydrates, and proteins. (a) What elements are present in fats? (b) What elements are present in carbohydrates? (c) What elements are present in proteins? 6. A compound that is present in many foods has the formula C44H86O8NP. To which class of molecules does this compound belong? Explain your answer. C CH2OH 7. A healthy diet must include some protein. (a) Assuming you had a way to measure the amount of each element in a sample of food, which element would you measure in order to tell whether the food contained protein? (b) The structures of three compounds are shown below. Based on your answer to part (a), which of the three compounds would you add to a food sample so that it would appear to contain more protein? Explain. (c) Which of the three compounds would already be present in a food sample that actually did contain protein? Explain. OH H N Nicotinic acid (niacin) Vitamin C O H3CO CH3 H3CO (CH2 CH3 CH C CH2)10H O Coenzyme Q 3. Name the four types of small biological molecules. Which three are capable of forming polymeric structures? What are the names of the polymeric structures that are formed? 4. To which of the four classes of biomolecules do the following compounds belong? (a) CH2OH O H H OH OH H H HO H NH CH3 C (a) (b) H O H2N NH2 O O H H CH2 CH2 CH NH⩽ 3 (d) C N C O 8. The structure of the compound urea is shown. Urea is a waste product of metabolism excreted by the kidneys into the urine. Why do doctors tell patients with kidney damage that they should consume a low-protein diet? H2N C NH2 OH COO⩾ 9. There are 20 different amino acids that occur in proteins (see Fig. 4-2). Each has the same basic structure with the exception of the R group, which is unique to each amino acid. What functional groups are present in all amino acids? 10. Draw the structure of the amino acid alanine. What is special about the central carbon atom of alanine? 11. The structures of the amino acids asparagine (Asn) and cysteine (Cys) are shown in Section 1-2. What functional group does Asn have that Cys does not? What functional group does Cys have that Asn does not? 12. Draw a dipeptide (a polypeptide containing two residues) that includes the two amino acids Asn and Cys shown in Section 1-2. When the peptide bond between the two residues is formed, which atoms are lost? Which functional groups are lost? Which new functional group is formed? O R H H OH (c) HS N CH2 O O⩾ CH2 NH2 Urea N P O⩾ O (b) ⩾O C O⩾ O O CH CH2 N N OH CH2OPO32⩾ ⩽H N 3 NH2 C H C O (c) O 13. The "straight-chain" structure of glucose is shown in Section 1-2. What functional groups are present in the glucose molecule? 21 14. Consider the monosaccharide fructose. (a) How does its molecular formula differ from that of glucose? (b) How does its structure differ from the structure of glucose? CH2OH (e) Several different types of lipid molecules assemble to form a membrane. 24. Does entropy increase or decrease in the following reactions in aqueous solution? (a) COO⩾ C O HO C H H C OH H C OH C C NH2 O Uracil H O ⩽ H⩽ C CH3 O ⩽ CO2(g) CH3 25. Which has the greater entropy, a polymeric molecule or a mixture of its constituent monomers? 26. How does the entropy change when glucose undergoes combustion? N O O CH2 (b) COO⩾ 15. The structures of the nitrogenous bases uracil and cytosine are shown below. How do their functional groups differ? N H C COO⩾ Fructose O O ⩽ CO2(g) CH3 CH2OH HN COO⩾ N H Cytosine 16. What are the structural components of the biological molecules called nucleotides? 17. Compare the solubilities in water of alanine, glucose, palmitate, and cholesterol, and explain your reasoning. 18. Cell membranes are largely hydrophobic structures. Which compound will pass through a membrane more easily, glucose or 2,4-dinitrophenol? Explain. C6H12O6 1 6 O2 n 6 CO2 1 6 H2O 27. A soccer coach keeps a couple of instant cold packs in her bag in case one of her players suffers a muscle injury. Instant cold packs are composed of a plastic bag containing a smaller water bag and solid ammonium nitrate. In order to activate the cold pack, the bag is kneaded until the smaller water bag breaks, which allows the released water to dissolve the ammonium nitrate. The equation for the dissolution of ammonium nitrate in water is shown below. How does the cold pack work? H 2O NH4NO3 (s) ¡ NH41 (aq) 1 NO2 3 (aq) DH 5 26.4 kJ?mol21 OH NO2 NO2 2,4-Dinitrophenol 28. Campers carry hot packs with them, especially when camping during the winter months or at high altitudes. Their design is similar to that described in Problem 27, except that calcium chloride is used in place of the ammonium nitrate. The equation for the dissolution of calcium chloride in water is shown below. How does the hot pack work? H2 O 19. What polymeric molecule forms a more regular structure, DNA or protein? Explain this observation in terms of the cellular roles of the two different molecules. CaCl2 (s) ¡ Ca21 (aq) 1 2Cl2 (aq) DH 5 281 kJ?mol21 29. For the reaction in which reactant A is converted to product B, tell whether this process is favorable at (a) 4⬚C and (b) 37⬚C. 20. What are the two major biological roles of polysaccharides? 21. Pancreatic amylase digests the glycosidic bonds that link glucose residues together in starch. Would you expect this enzyme to digest the glycosidic bonds in cellulose as well? Explain why or why not. A B H (kJ ? mol21) 54 60 S (J ? K21 ? mol21) 22 43 22. The complete digestion of starch in mammals yields 4 kilocalories per gram (see Problem 21). What is the energy yield for cellulose? 30. For a given reaction, the value of DH is 15 kJ ? mol21 and the value of DS is 51 J ? K21 ? mol21. Above what temperature will this reaction be spontaneous? 1-3 Energy and Metabolism 31. The hydrolysis of pyrophosphate at 258C is spontaneous. The enthalpy change for this reaction is 214.3 kJ ? mol21. What is the sign and the magnitude of DS for this reaction? 23. What is the sign of the entropy change for each of the following processes? (a) Water freezes. (b) Water evaporates. (c) Dry ice sublimes. (d) Sodium chloride dissolves in water. 32. Phosphoenolpyruvate donates a phosphate group to ADP to produce pyruvate and ATP. The ⌬G value for this reaction at 258C is 263 kJ ? mol21 and the value of DS is 190 J ? K21 ? mol21. What is the value of DH ? Is the reaction exothermic or endothermic? 22 33. Which of the following processes are spontaneous? (a) a reaction that occurs with any size decrease in enthalpy and any size increase in entropy (b) a reaction that occurs with a small increase in enthalpy and a large increase in entropy (c) a reaction that occurs with a large decrease in enthalpy and a small decrease in entropy (d) a reaction that occurs with any size increase in enthalpy and any size decrease in entropy 34. When a stretched rubber band is allowed to relax, it feels cooler. (a) Is the enthalpy change for this process positive or negative? (b) Since a stretched rubber band spontaneously relaxes, what can you conclude about the entropy change during relaxation? 35. Urea (NH2CONH2) dissolves readily in water; that is, this is a spontaneous process. The beaker containing the dissolved compound is cold to the touch. What conclusions can you make about the sign of the (a) enthalpy change and (b) entropy change for this process? 36. Phosphofructokinase catalyzes the transfer of a phosphate group (from ATP) to fructose-6-phosphate to produced fructose1,6-bisphosphate. The DH value for this reaction is 29.5 kJ ? mol21 and the ⌬G is 217.2 kJ ? mol21. (a) Is the reaction exothermic or endothermic? (b) What is the value of DS for the reaction? Does this reaction proceed with an increase or decrease in entropy? (c) Which component makes a greater contribution to the free energy value—the DH or DS value? Comment on the significance of this observation. 40. Identify the process described in the statements below as an oxidation or reduction process. (a) Monosaccharides are synthesized from carbon dioxide by plants during photosynthesis. (b) An animal eats the plant and breaks down the monosaccharide in order to obtain energy for cellular processes. 41. Given the following reactions, tell whether the reactant is being oxidized or reduced. Reactions may not be balanced. (a) CH3 (CH2)14 OH CH CH COO⩾ COO⩾ O CH2 O DG 5 ⩾30.5 kJ ? mol21 to ⩾O C H C OH H C B SH 42. For each of the reactions in Problem 41, tell whether an oxidizing agent or a reducing agent is needed to accomplish the reaction. 43. In some cells, lipids such as palmitate (shown in Section 1-2), rather than monosaccharides, serve as the primary metabolic fuel. (a) Consider the oxidation state of palmitate's carbon atoms and explain how it fits into a scheme such as the one shown in Figure 1-9. (b) On a per-carbon basis, which would make more free energy available for metabolic reactions: palmitate or glucose? 44. Which yields more free energy when completely oxidized, stearate or ␣-linolenate? H3C (CH2)16 COO⩾ Stearate H3C CH2 (CH CHCH2)3 (CH2)6 COO⩾ ␣-Linolenate 1-4 The Origin and Evolution of Life H O C H A O⩾ O (b) Write the equation for the overall conversion of GAP to 3PG. Is the coupled reaction favorable? 39. Place these molecules in order from the most oxidized to the most reduced. C CH2 CH NH⩽ 3 DG 5 218.8 kJ ? mol21 H 2 ⩽H3N CH ⩽ H2 CH2 (a) Is this reaction spontaneous? The reaction shown above is coupled to the following reaction in which 1,3BPG is converted to 3-phosphoglycerate (3PG): O OH O⩾ C S GAP 1 Pi 1 NAD1 ÷ 1,3BPG 1 NADH DG 5 16.7 kJ ? mol21 1,3BPG 1 ADP ÷ 3PG 1 ATP CH S (a)

Is this reaction favorable? Explain. (b) Suppose the synthesis of glucose-6-phosphate is coupled with the hydrolysis of ATP. Write the overall equation for the coupled process and calculate the ⌬G of the coupled reaction. Is the conversion of glucose to glucose-6-phosphate favorable under these conditions? Explain. 38. Glyceraldehyde-3-phosphate
(GAP) is converted 1,3-bisphosphoglycerate (1,3BPG) as shown below. COO⫺ CH2 glucose 1 phosphate ÷ glucose-6-phosphate 1 H2O DG 5 13.8 kJ?mol21 CoA COO⫺ CH 3N S O C (c) COO⫺ ⫹H C CH2 COO⫺ (d) 8 CH3 COO⫺ CH2 CH O⫺ C (b) COO⫺ 37. Glucose can be converted to glucose-6-phosphate: ATP 1 H2O ÷ ADP 1 phosphate O O C O 45. In
the 1920s, Oparin and Haldane independently suggested that the energy of lightning storms in the prebiotic world could have transformed gases in the early atmosphere to small organic molecules. In 1953, Miller and Urey conducted an experiment 23 demonstrating that this scenario is plausible. They subjected a mixture of water, methane,
ammonia, and hydrogen gas to electric discharge and refluxed the mixture so that any compounds formed would dissolve and accumulate in the water. At the end of a week, they analyzed the solution and found glycine, alanine, lactic acid, urea, and other amino acids and small organic acids. What was the significance of this experiment? Species A
Species B Species B TCGTCGAGTC TGGACTAGCC TGGACCAGCC 50. A portion of the evolutionary tree for a flu virus is shown here. Different strains are identified by an H followed by a number. 46. In order to give rise to more highly complex structures, what capabilities did the first biological molecules have to have? H15 47. Why is molecular
information so important for classifying and tracing the evolutionary relatedness of bacterial species but less important for vertebrate species? H7 48. The first theories to explain the similarities between bacteria and mitochondria or chloroplasts suggested that an early eukaryotic cell actually engulfed but failed to fully digest a free-living prokaryotic
cell. Why is such an event unlikely to account for the origin of mitochondria or chloroplasts? 49. Draw a simple evolutionary tree that shows the relationships between species A, B, and C based on the DNA sequences given here. H10 H3 H4 H14 (a) Identify two pairs of closely related flu strains. (b) Which strain(s) is(are) most closely related to strain
H3? SELECTED READINGS Koonin, E. V., The origin and early evolution of eukaryotes in the light of phylogenomics, Genome Biol. 11, 209 (2010). [Discusses several models for the origin of eukaryotic cells.] Koshland, D. E., Jr., The seven pillars of life, Science 295, 2215–2216 (2002). [Describes some of the essential attributes of all organisms,
including a DNA program, ability to mutate, compartmentalization, need for energy, ability to regenerate, adaptability, and seclusion.] Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. B., and Worm, B., How many species are there on earth and in the ocean? PLoS Biol 9(8): e1001127. doi:10.1371/journal.pbio.1001127 (2011). [Shows how statistical
analysis of databases can be used to estimate the total number of known and not-yet-described species.] Nee, S., More than meets the eye, Nature 429, 804–805 (2004). [A brief commentary about appreciating the metabolic diversity of microbial life.] Nisbet, E. G., and Sleep, N. H., The habitat and nature of early life, Nature 409, 1083–1091 (2001).
[Explains some of the hypotheses regarding the early earth and the origin of life, including the possibility that life originated at hydrothermal vents.] Tinoco, I., Jr., Sauer, K., Wang, J. C., and Puglisi, J., Physical Chemistry: Principles and Applications in Biological Sciences (4th ed.), Chapters 2–5, Prentice Hall (2002). [This and other physical chemistry
textbooks present the basic equations of thermodynamics.] chapter 2 AQUEOUS CHEMISTRY WHY do so many substances dissolve in water? We call water the medium of life because water surrounds biological molecules and helps determine their shapes and chemical reactivity. But why is this so? In this chapter we will see how water interacts with
other substances, such as the green dye shown here—either through electrostatic or hydrophobic effects—in ways that are not possible for other types of solvents. [stuartbur/iStockphoto] THIS CHAPTER IN CONTEXT Do You Remember? Part 1 Foundations 2 Aqueous Chemistry Biological molecules are composed of a subset of all possible elements
and functional groups (Section 1-2). The free energy of a system is determined by its enthalpy and entropy (Section 1-3). Part 2 Molecular Structure and Function Part 3 Metabolism Part 4 Genetic Information 24 Living organisms can be found in virtually every portion of the earth that contains liquid water. In the polar ice caps, prokaryotes and small
eukaryotes survive in the spaces between ice crystals (Fig. 2-1). The hot waters near hydrothermal vents host the prokaryote Pyrolobus fumarii, which grows best at a temperature of 1058C and can tolerate 1138C. Living organisms have even been discovered several kilometers below the earth's surface, but only where water is present. Water is a
fundamental requirement for life, so it is important to understand the structural and chemical properties of water. Not only are most biological molecules surrounded by water, but their molecular structure is in part governed by how their component groups interact with water. And water plays a role in how these molecules assemble to form larger
structures or undergo chemical transformation. In fact, water itself—or its H1 and OH2 constituents—participates directly in many biochemical processes. Therefore, an examination of water is a logical prelude to exploring the structures and functions of biomolecules in the following chapters. Figure 2-1 Methanococcoides burtonii. These bacterial
cells from an Antarctic lake survive at temperatures as low as 22.58C. [M. Rohde, GBF/Science Photo Library/Photo Researchers.] 2-1 Water Molecules Form Hydrogen Bonds What is the nature of the substance that accounts for about 70% of the mass of most organisms? The human body, for example, is about 60% by weight water, most of it in the
extracellular fluid (the fluid surrounding cells) and inside cells: Intracellular water (40%) Non-water (40%) Extracellular water (15%) Water in circulatory system (5%) In an individual H2O molecule, the central oxygen atom forms covalent bonds with two hydrogen atoms, leaving two unshared pairs of electrons. The molecule therefore has
approximately tetrahedral geometry, with the oxygen atom at the center of the tetrahedron, the hydrogen atoms at two of the four corners, and electrons at the other two corners (Fig. 2-2). As a result of this electronic arrangement, the water molecule is polar ; that is, it has an uneven distribution of charge. The oxygen atom bears a partial negative
charge (indicated by the symbol d2), and each hydrogen atom bears a partial positive charge (indicated by the symbol d1): δ+ KEY CONCEPTS • The polar water molecule forms hydrogen bonds with other molecules. • Noncovalent forces, including hydrogen bonds, ionic interactions, and van der Waals forces, act on biological molecules. • Water
dissolves both ionic and polar substances. δ+ H O H Figure 2-2 Electronic structure of the water molecule. Four electron δ− This polarity is the key to many of water's unique physical properties. orbitals, in an approximately tetrahedral arrangement, surround the central oxygen. Two orbitals participate in bonding to hydrogen (gray), and two contain
unshared electron pairs. 26 Ch 2 Aqueous Chemistry Neighboring water molecules tend to orient themselves so that each partially positive hydrogen is aligned with a partially negative oxygen: δ− δ+ Figure 2-3 Structure of ice. Each water molecule acts as a donor for two hydrogen bonds and an acceptor for two hydrogen bonds, thereby interacting
with four other water molecules in the crystal. (Only two layers of water molecules are shown here.) ? Identify a hydrogen bond donor and acceptor in this structure. This interaction, shaded yellow here, is known as a hydrogen bond. Traditionally shown as a simple electrostatic attraction between oppositely charged particles, the hydrogen bond is
now known to have some covalent character. This means that the bond has directionality, or a preferred orientation. Each water molecule can potentially participate in four hydrogen bonds, since it has two hydrogen atoms to "donate" to a hydrogen bond and two pairs of unshared electrons that can "accept" a hydrogen bond. In ice, a crystalline form
of water, each water molecule does indeed form hydrogen bonds with four other water molecules (Fig. 2-3). This regular, lattice-like structure breaks down when the ice melts. In liquid water, each molecule can potentially form hydrogen bonds with up to four other water molecules, but each bond has a lifetime of only about 10212 s. As a result, the
structure of water is continually flickering as water molecules rotate, bend, and reorient themselves. Theoretical calculations and spectroscopic data suggest that water molecules participate in only two strong hydrogen bonds, one as a donor and one as an acceptor, generating transient hydrogen-bonded clusters such as the sixmembered ring shown
here: Because of its ability to form hydrogen bonds, water is highly cohesive. This accounts for its high surface tension, which allows certain insects to walk on water (Fig. 2-4). The cohesiveness of water molecules also explains why water remains a liquid, whereas molecules of similar size, such as CH4 and H2S, are gases at room temperature
(258C). At the same time, water is less dense than other liquids because hydrogen bonding demands that individual molecules not just approach each other but interact with a certain orientation. This geometrical constraint also explains why ice floats; for other materials, the solid is denser than the liquid. Hydrogen bonds are one type of electrostatic
force Figure 2-4 A water strider supported by the surface tension of water. [Hermann Eisenbeiss/Photo Research, Inc.] Powerful covalent bonds define basic molecular constitutions, but much weaker noncovalent bonds, including hydrogen bonds, govern the final three-dimensional shapes of molecules and how they interact with each other. For
example, about 460 kJ ? mol21 (110 kcal ? mol21) of energy is required to break a covalent OOH bond. But a hydrogen bond in water has a strength of only about 20 kJ ? mol21 (4.8 kcal ? mol21). Other noncovalent interactions are weaker still. Among the noncovalent interactions that occur in biological molecules are electrostatic interactions
between charged groups such as carboxylate (OCOO2) and amino (ONH1 3) groups. These ionic interactions are intermediate in strength to covalent bonds and hydrogen bonds (Fig. 2-5). (a) O Covalent bond H 27 Water Molecules Form Hydrogen Bonds 400 350 Bond strength (kJ . mol−1) Hydrogen bonds, despite their partial covalent nature, are
classified as a type of electrostatic interaction. At about 1.8 Å, they are longer and hence weaker than a covalent OOH bond (which is about 1 Å long). However, a completely noninteracting O and H would approach no closer than about 2.7 Å, which is the sum of their van der Waals radii (the van der Waals radius of an isolated atom is the distance
from its nucleus to its effective electronic surface). 300 250 200 150 100 50 1Å 0 Covalent bond (b) O H Hydrogen bond Ionic Hydrogen interaction bond van der Waals interaction Figure 2-5 Relative strengths of bonds in biological molecules. 1.8 Å (c) O No bond H 2.7 Å Hydrogen bonds usually involve NOH, OOH, and SOH groups as hydrogen
donors and the electronegative N, O, or S atoms as hydrogen acceptors (electronegativity is a measure of an atom's affinity for electrons; Table 2-1). Water, therefore, can form hydrogen bonds not just with other water molecules but with a wide variety of other compounds that bear N-, O-, or S-containing functional groups. H H O H H O R O R H N H
R Water–alcohol Water–amine Likewise, these functional groups can form hydrogen bonds among themselves. For example, the complementarity of bases in DNA and RNA is determined by their ability to form hydrogen bonds with each other: Guanine Cytosine Other electrostatic interactions occur between particles that are polar but not actually
charged, for example, two carbonyl groups: C O C O TABLE 2-1 Electronegativities of Some Elements Element Electronegativity C F H N O 2.55 3.98 2.20 3.04 3.44 28 Ch 2 Aqueous Chemistry These forces, called van der Waals interactions, are usually weaker than hydrogen bonds. The interaction shown on the previous page, between two
strongly polar groups, is known as a dipole–dipole interaction and has a strength of about 9 kJ ? mol21. Very weak van der Waals interactions, called London dispersion forces, occur between nonpolar molecules as a result of small fluctuations in their distribution of electrons that create a temporary separation of charge. Nonpolar groups such as
methyl groups can therefore experience a small attractive force, in this case about 0.3 kJ ? mol21: H C H of small forces. Just as the fictional giant Gulliver was restrained by many small tethers at the hands of the tiny Lilliputians, the structures of macromolecules are constrained by the effects of many weak noncovalent interactions. [Hulton
Archive/Getty Images.] H H C H H Figure 2-6 The cumulative effect Not surprisingly, these forces act only when the groups are very close, and their strength quickly falls off as the groups draw apart. If the groups approach too closely, however, their van der Waals radii collide and a strong repulsive force overcomes the attractive force. Although
hydrogen bonds and van der Waals interactions are individually weak, biological molecules usually contain multiple groups capable of participating in these intermolecular interactions, so their cumulative effect can be significant (Fig. 2-6). Drug molecules are typically designed to optimize the weak interactions that govern their therapeutic activity
(Box 2-A). BOX 2-A BIOCHEMISTRY NOTE Why Do Some Drugs Contain Fluorine? As mentioned in Section 1-2, the most abundant elements in biological molecules are H, C, N, O, P, and S. Fluorine only rarely appears in naturally occurring organic compounds. Why, then, do about one-quarter of all drug molecules, including the widely prescribed
Prozac (fluoxetine, an antidepressant; Box 9-C), fluorouracil (an anticancer agent; Section 7-3), and Ciprofloxacin (an antibacterial agent; Section 20-1), contain F? O H N CH3 F3C Prozac (Fluoxetine) In designing an effective drug, pharmaceutical scientists often intentionally introduce F in order to alter the drug's chemical or biological properties
without significantly altering its shape. Fluorine can take the place of hydrogen in a chemical structure, but with its high electronegativity (see Table 2-1), F behaves much more like O than H. Consequently, transforming a relatively inert COH group into an electron-withdrawing C—F group can decrease the basicity of nearby amino groups (see
Section 2-3). Fewer positive charges in a drug allow it to more easily pass through membranes to enter cells and exert its biological effect. In addition, the polar COF bond can participate in hydrogen bonding (COF · · · HOC) or other dipole–dipole interactions (such as COF ? ? ? CPO), potentially augmenting the intermolecular attraction between a
drug and its target molecule in the body. Better binding usually means that the drug will be effective at lower concentrations and will have fewer side effects. Question: Identify the hydrogen-bonding groups in Prozac. Water dissolves many compounds 29 The answer to the question we asked at the start of the chapter is that, unlike most other solvent
molecules, water molecules are able to form hydrogen bonds and participate in other electrostatic interactions with a wide variety of compounds. Water has a relatively high dielectric constant, which is a measure of a solvent's ability to diminish the electrostatic attractions between dissolved ions (Table 2-2). The higher the dielectric constant of the
solvent, the less able the ions are to associate with each other. The polar water molecules surround ions (for example, the Na1 and Cl2 ions from the salt NaCl) by aligning their partial charges with the oppositely charged ions. Because the interactions between the polar water molecules and the ions are stronger than the attractive forces between the
Na1 and Cl2 ions, the salt dissolves (the dissolved particle is called a solute). Each solute ion surrounded by water molecules (shown at right) is said to be solvated (or hydrated, to indicate that the solvent is water). Biological molecules that bear polar or ionic functional groups are also readily solubilized, in this case because the groups can form
hydrogen bonds with the solvent water molecules. Glucose, for example, with its six hydrogen-bonding oxygens, is highly soluble in water: WHY do so many substances dissolve in water? Na+ Cl– The concentration of glucose in human blood is about 5 mM. In a solution of 5 mM glucose in water, there are about 10,000 water molecules for every
glucose molecule (the water molecules are present at a concentration of about 55.5 M). However, biological molecules are never found alone in such dilute conditions in vivo, because a large number of small molecules, large polymers, and macromolecular aggregates collectively form a solution that is more like a hearty stew than a thin, watery soup
(Fig. 2-7). Inside a cell, the spaces between molecules may be only a few Å wide, enough room for only two water molecules to fit. This allows solute molecules, each with a coating of properly oriented water molecules, to slide past each other. This thin coating, or shell, of water may be enough to keep molecules from coming into TABLE 2-2 Dielectric
Constants for Some Solvents at Room Temperature Solvent Dielectric Constant Formamide (HCONH2) Water Methanol (CH3OH) Ethanol (CH3CH2OH) 1-Propanol (CH3CH2CH2OH) 1-Butanol (CH3CH2CH2CH2OH) Benzene (C6H6) ? Compare the hydrogen-bonding ability of these solvents. 109 80 33 25 20 18 2 Figure 2-7 Portion of a Dictyostelium
cell visualized by cryoelectron tomography. In this technique, the cells are rapidly frozen so that they retain their fine structure, and two-dimensional electron micrographs taken from different angles are merged to re-create a threedimensional image. The red structures are filaments of the protein actin, ribosomes and other macromolecular
complexes are colored green, and membranes are blue. Small molecules (not visible) fill the spaces between these larger cell components. [Courtesy Wolfgang Baumeister, Max Planck Institute for Biochemistry.] 30 Ch 2 Aqueous Chemistry van der Waals contact (van der Waals interactions are weak but attractive), thereby helping maintain the cell's
contents in a crowded but fluid state. CONCEPT REVIEW • • • • Why is a water molecule polar? What is a hydrogen bond, and why does it form between water molecules? Describe the structure of liquid water. Describe the nature and relative strength of ionic interactions, hydrogen bonds, and van der Waals interactions. • What happens when an
ionic substance dissolves? • Explain why water is a more effective solvent than ammonia or methanol. 2-2 The Hydrophobic Effect KEY CONCEPTS • The hydrophobic effect, which is driven by entropy, excludes nonpolar substances from water. • Amphiphilic molecules form micelles or bilayers. Glucose and other readily hydrated substances are said
to be hydrophilic (waterloving). In contrast, a compound such as dodecane (a C12 alkane), which lacks polar groups, is relatively insoluble in water and is said to be hydrophobic (water-fearing). Although pure hydrocarbons are rare in biological systems, many biological molecules contain hydrocarbon-like portions that are insoluble in water. When a
nonpolar substance such as vegetable oil (which consists of hydrocarbonlike molecules) is added to water, it does not dissolve but forms a separate phase. In order for the water and oil to mix, free energy must be added to the system (for example, by stirring vigorously or applying heat). Why is it thermodynamically unfavorable to dissolve a
hydrophobic substance in water? One possibility is that enthalpy is required to break the hydrogen bonds among solvent water molecules in order to create a "hole" into which a nonpolar molecule can fit. Experimental measurements, however, show that the free energy barrier (DG) to the solvation process depends much more on the entropy term
(DS) than on the enthalpy term (DH; recall from Chapter 1 that DG 5 DH 2 T DS; Equation 1-2). This is because when a hydrophobic molecule is hydrated, it becomes surrounded by a layer of water molecules that cannot participate in normal hydrogen bonding but instead must align themselves so that their polar ends are not oriented toward the
nonpolar solute. This constraint on the structure of water represents a loss of entropy in the system, because now the highly mobile water molecules have lost some of their freedom to rapidly form, break, and re-form hydrogen bonds with other water molecules (Fig. 2-8). The loss Nonpolar molecule Water Layer of constrained water molecules Figure
2-8 Hydration of a nonpolar molecule. When a nonpolar molecule (green) is added to water, the system loses entropy because the water molecules surrounding the nonpolar solute (orange) lose their freedom to form hydrogen bonds. The loss of entropy is a property of the entire system, not just the water molecules nearest the solute, because these
molecules are continually changing places with water molecules from the rest of the solution. The loss of entropy presents a thermodynamic barrier to the hydration of a nonpolar solute. Figure 2-9 Aggregation of nonpolar molecules in water. (a) The individual (a) (b) of entropy is not due to the formation of a frozen "cage" of water molecules around
the nonpolar solute, as commonly pictured, because in liquid water, the solvent molecules are in constant motion. When a large number of nonpolar molecules are introduced into a sample of water, they do not disperse and become individually hydrated, each surrounded by a layer of water molecules. Instead, the nonpolar molecules tend to clump
together, removing themselves from contact with water molecules. (This explains why small oil droplets coalesce into one large oily phase.) Although the entropy of the nonpolar molecules is thereby reduced, this thermodynamically unfavorable event is more than offset by the increase in the entropy of the water molecules, which regain their ability
to interact freely with other water molecules (Fig. 2-9). The exclusion of nonpolar substances from an aqueous solution is known as the hydrophobic effect. It is a powerful force in biochemical systems, even though it is not a bond or an attractive interaction in the conventional sense. The nonpolar molecules do not experience any additional attractive
force among themselves; they aggregate only because they are driven out of the aqueous phase by the unfavorable entropy cost of individually hydrating them. The hydrophobic effect governs the structures and functions of many biological molecules. For example, the polypeptide chain of a protein folds into a globular mass so that its hydrophobic
groups are in the interior, away from the solvent, and its polar groups are on the exterior, where they can interact with water. Similarly, the structure of the lipid membrane that surrounds all cells is maintained by the hydrophobic effect acting on the lipids. hydration of dispersed nonpolar molecules (green) decreases the entropy of the system
because the hydrating water molecules (orange) are not as free to form hydrogen bonds. (b) Aggregation of the nonpolar molecules increases the entropy of the system, since the number of water molecules required to hydrate the aggregated solutes is less than the number of water molecules required to hydrate the dispersed solute molecules. This
increase in entropy accounts for the spontaneous aggregation of nonpolar substances in water. ? Explain why it is incorrect to describe the behavior shown in part (b) in terms of "hydrophobic bonds." Amphiphilic molecules experience both hydrophilic interactions and the hydrophobic effect Polar head group Consider a molecule such as the fatty acid
palmitate: Nonpolar tail The hydrocarbon "tail" of the molecule (on the right) is nonpolar, while its carboxylate "head" (on the left) is strongly polar. Molecules such as this one, which have both hydrophobic and hydrophilic portions, are said to be amphiphilic or amphipathic. What happens when amphiphilic molecules are added to water? In general,
the polar groups of amphiphiles orient themselves toward the solvent molecules and are therefore hydrated, while the nonpolar groups tend to aggregate due to the hydrophobic effect. As a result, the amphiphiles may form a spherical micelle, a particle with a solvated surface and a hydrophobic core (Fig. 2-10). Depending in part on the relative sizes
of the hydrophilic and hydrophobic portions of the amphiphiles, the molecules may form a sheet rather than a spherical micelle. The amphiphilic lipids that provide the structural basis of biological membranes form two-layered sheets called bilayers, in which a hydrophobic layer is Figure 2-10 A micelle formed by amphiphilic molecules. The
hydrophobic tails of the molecules aggregate, out of contact with water, due to the hydrophobic effect. The polar head groups are exposed to and can interact with the solvent water molecules. sandwiched between hydrated polar surfaces (Fig. 2-11). The structures of biological membranes are discussed in more detail in Chapter 8. The formation of
micelles or bilayers is thermodynamically favored because the hydrogen-bonding capacity of the polar head groups is satisfied through interactions with solvent water molecules, and the nonpolar tails are sequestered from the solvent. 32 Ch 2 Aqueous Chemistry Polar head group Nonpolar tails The hydrophobic core of a lipid bilayer is a barrier to
diffusion To eliminate its solvent-exposed edges, a lipid bilayer tends to close up to form a vesicle, shown cut in half: Figure 2-11 A lipid bilayer. The amphiphilic lipid molecules form two layers so that their polar head groups are exposed to the solvent while their hydrophobic tails are sequestered in the interior of the bilayer, away from water. The
likelihood of amphiphilic molecules forming a bilayer rather than a micelle depends in part on the sizes and nature of the hydrophobic and hydrophilic groups. One-tailed lipids tend to form micelles (see Fig. 2-10), and two-tailed lipids tend to form bilayers. ? Indicate where a sodium ion and a benzene molecule would be located. Many of the
subcellular compartments (organelles) in eukaryotic cells have a similar structure. When the vesicle forms, it traps a volume of the aqueous solution. Polar solutes in the enclosed compartment tend to remain there because they cannot easily pass through the hydrophobic interior of the bilayer. The energetic cost of transferring a hydrated polar group
through the nonpolar lipid tails is too great. (In contrast, small nonpolar molecules such as O2 can pass through the bilayer relatively easily.) Normally, substances that are present at high concentrations tend to diffuse to regions of lower concentration (this movement "down" a concentration gradient is a spontaneous process driven by the increase in
entropy of the solute molecules). A barrier such as a bilayer can prevent this diffusion (Fig. 2-12). This helps explain why cells, which are universally enclosed by a membrane, can maintain their specific concentrations of ions, small molecules, and biopolymers even when the external concentrations of these substances are quite different (Fig. 2-13).
The solute composition of intracellular compartments and other biological fluids is carefully regulated. Not surprisingly, organisms spend a considerable amount of metabolic energy to maintain the proper concentrations of water and salts, and losses of one or the other must be compensated (Box 2-B). INTRACELLULAR EXTRACELLULAR (b) Figure
2-12 A bilayer prevents the diffusion of polar substances. (a) Solutes spontaneously diffuse from a region of high concentration to a region of low concentration. (b) A lipid barrier, which presents a thermodynamic barrier to the passage of polar substances, prevents the diffusion of polar substances out of the inner compartment (it also prevents the
inward diffusion of polar substances from the external solution). 160 Concentration (mM) (a) Concentration (mM) 160 120 80 40 0 Na+ K+ Cl − 120 80 40 0 Na+ K+ Cl − Figure 2-13 Ionic composition of intracellular and extracellular fluid. Human cells contain much higher concentrations of potassium than of sodium or chloride; the opposite is true
of the fluid outside the cell. The cell membrane helps maintain the concentration differences. BOX 2-B BIOCHEMISTRY NOTE 33 Acid–Base Chemistry Sweat, Exercise, and Sports Drinks Animals, including humans, generate heat, even at rest, due to their metabolic activity. Some of this heat is lost to the environment by radiation, convection,
conduction, and—in terrestrial animals—the vaporization of water. Evaporation has a significant cooling effect because about 2.5 kJ of heat is given up for every gram (mL) of water lost. In humans and certain other animals, an increase in skin temperature triggers the activity of sweat glands, which secrete a solution containing (in humans) about 50
mM Na1, 5 mM K1, and 45 mM Cl2. The body is cooled as the sweat evaporates from its surface. The evaporation of water accounts for a small portion of a resting body's heat loss, but sweating is the main mechanism for dissipating heat generated when the body is highly active. During vigorous exercise or exertion at high ambient temperatures, the
body may experience a fluid loss of up to 2 L per hour. Athletic training not only improves the performance of the muscles and cardiopulmonary system, it also increases the capacity for sweating so that the athlete begins to sweat at a lower skin temperature and loses less salt in the secretions of the sweat glands. But regardless of training, a fluid
loss representing more than 2% of the body's weight may impair cardiovascular function. In fact, "heat exhaustion" in humans is usually due to dehydration rather than an actual increase in body temperature. Numerous studies have concluded that athletes seldom drink enough before or during exercise. Ideally, fluid intake should match the losses
due to sweat, and the rate of intake should keep pace with the rate of sweating. So what should the conscientious athlete drink? For activities lasting less than about 90 minutes, especially when periods of high intensity alternate with brief periods of rest, water alone is sufficient. Commercial sports drinks containing carbohydrates can replace the
water lost as sweat and also provide a source of energy. However, this carbohydrate boost may be an advantage only during prolonged sustained activity, such as during a marathon, when the body's own carbohydrate stores are depleted. A marathon runner or a manual laborer in the hot sun might benefit from the salt found in sports drinks, but most
athletes don't need the supplemental salt (although it does make the carbohydrate solution more palatable). A normal diet usually contains enough Na1 and Cl2 to offset the losses in sweat. Question: Compare the ion concentrations of sweat and extracellular fluid. CONCEPT REVIEW • Why do polar molecules dissolve more easily than nonpolar
substances in water? • What does entropy have to do with the solubility of nonpolar substances in water? • Can a molecule be both hydrophilic and hydrophobic? • Why is a lipid bilayer a barrier to the diffusion of polar molecules? 2-3 Acid–Base Chemistry Water is not merely an inert medium for biochemical processes; it is an active participant. Its
chemical reactivity in biological systems is in part a result of its ability to ionize. This can be expressed in terms of a chemical equilibrium: H2O Δ H 1 1 OH 2 The products of water's dissociation are a hydrogen ion or proton (H1) and a hydroxide ion (OH2). KEY CONCEPTS • Water ionizes to form H1 and OH2. • An acid's pK value describes its
tendency to ionize. • The pH of a solution of acid depends on the pK and the concentrations of the acid and its conjugate base. Aqueous solutions do not actually contain lone protons. Instead, the H1 can be visualized as combining with a water molecule to produce a hydronium ion (H3O1): 34 Ch 2 Aqueous Chemistry H H O H However, the H is
somewhat delocalized, so it probably exists as part of a larger, fleeting structure such as 1 + Figure 2-14 Proton jumping. A proton associated with one water molecule (as a hydronium ion, at left) appears to jump rapidly through a network of hydrogen-bonded water molecules. H H O H O H H H H O O or H H H H O H H O H Because a proton does
not remain associated with a single water molecule, it appears to be relayed through a hydrogen-bonded network of water molecules (Fig. 2-14). This rapid proton jumping means that the effective mobility of H1 in water is much greater than the mobility of other ions that must physically diffuse among water molecules. Consequently, acid–base
reactions are among the fastest biochemical reactions. [H1] and [OH2] are inversely related Pure water exhibits only a slight tendency to ionize, so the resulting concentrations of H1 and OH2 are actually quite small. According to the law of mass action, the ionization of water can be described by a dissociation constant, K, which is equivalent to the
concentrations of the reaction products divided by the concentration of un-ionized water: K5 [H 1][OH 2] [H2O] [2-1] The square brackets represent the molar concentrations of the indicated species. Because the concentration of H2O (55.5 M) is so much greater than [H1] or [OH2], it is considered to be constant, and K is redefined as K w, the
ionization constant of water: K w 5 K [H2O] 5 [H 1][OH 2] 214 1 [2-2] 2 1 K w is 10 at 258C. In a sample of pure water, [H] 5 [OH], so [H] and [OH2] must both be equal to 1027 M: K w 5 10 214 5 [H 1][OH 2] 5 (10 27 M)(10 27 M) ⫺ ⫺ 214 [2-3] Since the product of [H] and [OH] in any solution must be equal to 10 , a hydrogen ion concentration
greater than 1027 M is balanced by a hydroxide ion concentration less than 1027 M (Fig. 2-15). 1 Figure 2-15 Relationship between 1 2 1 [H] and [OH]. The product of [H] and [OH2] is K w, which is equal to 10214. Consequently, when [H1] is greater than 1027 M, [OH2] is less than 1027 M, and vice versa. [H+] (M) 1 10 −2 10 −2 10 −4 10 −4 10
−6 10 −6 10 −8 10 −8 10 −10 10 −10 10 −12 10 −12 10 −14 10 −14 [OH−] (M) A solution in which [H1] 5 [OH2] 5 1027 M is said to be neutral; a solution with [H1] . 1027 M ([OH2] , 1027 M) is acidic; and a solution with [H1] , 1027 M ([OH2] . 1027 M) is basic. To more easily describe such solutions, the hydrogen ion concentration is expressed as
a pH: pH 5 2log[H 1] 35 Acid–Base Chemistry [2-4] Accordingly, a neutral solution has a pH of 7, an acidic solution has a pH , 7, and a basic solution has a pH . 7 (Fig. 2-16). Note that because the pH scale is logarithmic, a difference of one pH unit is equivalent to a 10-fold difference in [H ⫺]. The so-called physiological pH, the normal pH of human
blood, is a near-neutral 7.4. The pH values of some other body fluids are listed in Table 2-3. Acid–base chemistry is also a concern outside the laboratory (Box 2-C). [H+] (M) 14 10 −14 13 10 −13 12 10 −12 11 10 −11 10 10 −10 9 10 −9 8 10 −8 7 10 −7 6 10 −6 5 10 −5 4 10 −4 3 10 −3 2 10 −2 1 10 −1 0 1 Basic Neutral BOX 2-C pH BIOCHEMISTRY
NOTE Acidic Atmospheric CO2 and Ocean Acidification The human-generated increase in atmospheric carbon dioxide that is contributing to global warming is also impacting the chemistry of the world's oceans. Atmospheric CO2 dissolves in water and reacts with it to generate carbonic acid. The acid immediately dissociates to form protons (H1) and
bicarbonate (HCO2 3): CO2 1 H2O Δ H2CO3 Δ H 1 1 HCO32 The addition of hydrogen ions from CO2-derived carbonic acid therefore leads to a decrease in the pH. Currently, the earth's oceans are slightly basic, with a pH of approximately 8.0. It has been estimated that over the next 100 years, the ocean pH will drop to about 7.8. Although the
oceans act as a CO2 "sink" that helps mitigate the increase in atmospheric CO2, the increase in acidity in the marine environment represents an enormous challenge to organisms that must adapt to the new conditions. Many marine organisms, including mollusks, many corals, and some plankton, use dissolved carbonate ions (CO22 3) to construct
protective shells of calcium carbonate (CaCO3). However, carbonate ions can combine with H1 to form bicarbonate: CO22 1 H 1 Δ HCO32 3 Consequently, the increase in ocean acidity could decrease the availability of carbonate and thereby slow the growth of shell-building organisms. This not only would affect the availability of shellfish for human
consumption but also would impact huge numbers of unicellular organisms at the base of the marine food chain. It is possible that acidification of the oceans could also dissolve existing calcium carbonate–based materials, such as coral reefs: CaCO3 1 H 1 Δ HCO32 1 Ca21 This could have disastrous consequences for these species-rich ecosystems.
Question: Paradoxically, some marine organisms appear to benefit from increased atmospheric CO2. Write an equation that describes how increased bicarbonate concentrations in seawater could promote shell growth. Figure 2-16 Relationship between pH and [H1]. Because pH is equal to –log [H1], the greater the [H1], the lower the pH. A solution
with a pH of 7 is neutral, a solution with a pH , 7 is acidic, and a solution with a pH . 7 is basic. ? What is the difference in H1 concentration between a solution at pH 4 and a solution at pH 8? TABLE 2-3 pH Values of Some Biological Fluids Fluid Pancreatic juice Blood Saliva Urine Gastric juice pH 7.8–8.0 7.4 6.4–7.0 5.0–8.0 1.5–3.0 36 Ch 2 Aqueous
Chemistry The pH of a solution can be altered The pH of a sample of water can be changed by adding a substance that affects the existing balance between [H1] and [OH2]. Adding an acid increases the concentration of [H1] and decreases the pH; adding a base has the opposite effect. Biochemists define an acid as a substance that can donate a
proton and a base as a substance that can accept a proton. For example, adding hydrochloric acid (HCl) to a sample of water increases the hydrogen ion concentration ([H1] or [H3O1]) because the HCl donates a proton to water: HCl 1 H2O n H3O 1 1 Cl 2 Note that in this reaction, H2O acts as a base that accepts a proton from the added acid.
Similarly, adding the base sodium hydroxide (NaOH) increases the pH (decreases [H1]) by introducing hydroxide ions that can recombine with existing hydrogen ions: NaOH 1 H3O 1 n Na 1 1 2 H2O In this reaction, H3O1 is the acid that donates a proton to the added base. The final pH of the solution depends on how much H1 (for example, from HCl)
has been introduced or how much H1 has been removed from the solution by its reaction with a base (for example, the OH2 ion of NaOH). Substances such as HCl and NaOH are known as "strong" acids and bases because they ionize completely in water. The Na1 and Cl2 ions are called spectator ions and do not affect the pH. Calculating the pH of a
solution of strong acid or base is straightforward (see Sample Calculation 2-1). SAMPLE CALCULATION 2-1 PROBLEM SOLUTION Calculate the pH of 1 L of water to which is added (a) 10 mL of 5.0 M HCl or (b) 10 mL of 5.0 M NaOH. (0.01 L)(5.0 M) 5 0.050 M 1.01 L Since HCl dissociates completely, the added [H1] is equal to [HCl], or 0.050 M (the
existing hydrogen ion concentration, 1027 M, can be ignored because it is much smaller). (a) The final concentration of HCl is pH 5 2log[H 1] 5 2log 0.050 5 1.3 (b) The final concentration of NaOH is 0.050 M. Since NaOH dissociates completely, the added [OH2] is 0.050 M. Use Equation 2-2 to calculate [H1]. Kw 5 [H 1] 5 5 5 pH 5 5 5 PRACTICE
PROBLEMS 10 214 5 [H 1][OH 2] 10 214y[OH 2] 10 214y(0.050 M) 2.0 3 10 213 M 2log[H 1] 2log(2.0 3 10 213) 12.7 1. Calculate the pH of 500 mL of water after the addition of 50 mL of 25 mM HCl. 2. Calculate the pH of 250 mL of water after mixing with 250 mL of 5 mM NaOH. 37 Acid–Base Chemistry A pK value describes an acid's tendency to
ionize Most biologically relevant acids and bases, unlike HCl and NaOH, do not dissociate completely when added to water. In other words, proton transfer to or from water is not complete. Therefore, the final concentrations of the acidic and basic species (including water itself) must be expressed in terms of an equilibrium. For example, acetic acid
partially ionizes, or donates only some of its protons to water: CH3COOH 1 H2O Δ CH3COO 2 1 H3O 1 The equilibrium constant for this reaction takes the form K5 [CH3COO 2][H3O 1] [CH3COOH][H2O] [2-5] Because the concentration of H2O is much higher than the other concentrations, it is considered constant and is incorporated into the value
of K, which is then formally known as K a, the acid dissociation constant: K a 5 K [H2O] 5 [CH3COO 2][H 1] [CH3COOH] [2-6] The acid dissociation constant for acetic acid is 1.74 3 1025. The larger the value of K a, the more likely the acid is to ionize; that is, the greater its tendency to donate a proton to water. The smaller the value of K a, the less
likely the compound is to donate a proton. Acid dissociation constants, like hydrogen ion concentrations, are often very small numbers. Therefore, it is convenient to transform the K a to a pK value as follows: pK 5 2log K a [2-7] pK 5 2log(1.74 3 10 25) 5 4.76 [2-8] For acetic acid, The larger an acid's Ka, the smaller its pK and the greater its
"strength" as an acid. Consider an acid such as the ammonium ion, NH14 : NH41 Δ NH3 1 H 1 Its K a is 5.62 3 10210, which corresponds to a pK of 9.25. This indicates that the ammonium ion is a relatively weak acid, a compound that tends not to donate a proton. On the other hand, ammonia (NH3), which is the conjugate base of the acid NH14 ,
readily accepts a proton. The pK values of some compounds are listed in Table 2-4. A polyprotic acid, a compound with more than one acidic hydrogen, has a pK value for each dissociation (called pK1, pK2, etc.). The first proton dissociates with the lowest pK value. Subsequent protons are less likely to dissociate and so have higher pK values. The pH
of a solution of acid is related to the pK When an acid (represented as the proton donor HA) is added to water, the final hydrogen ion concentration of the solution depends on the acid's tendency to ionize: HA Δ A 2 1 H 1 38 Ch 2 Aqueous Chemistry TABLE 2.4 pK Values of Some Acids Name Formulaa pK Trifluoroacetic acid Phosphoric acid Formic
acid Succinic acid Acetic acid Succinate Thiophenol Phosphate N-(2-acetamido)-2-aminoethanesulfonic acid (ACES) CF3COOH H3PO4 HCOOH HOOCCH2CH2COOH CH3COOH HOOCCH2CH2COO2 C6H5SH H2PO42 0.18 2.15b 3.75 4.21b 4.76 5.64c 6.60 6.82c 1 H2NCOCH2NH2CH2CH2SO32 6.90 NH Imidazolium ion 7.00 N H p-Nitrophenol HO
N-2-hydroxyethylpiperazineN9-2-ethanesulfonic acid (HEPES) HOCH2CH2HN Glycinamide Tris(hydroxymethyl)aminomethane (Tris) Boric acid Ammonium ion Phenol Methylammonium ion Phosphate a NO2 1 NCH2CH2SO 3 H3NCH2CONH2 7.24 7.55 8.20 1 (HOCH2) 3CNH3 H3BO3 NH14 C6H5OH CH3NH 13 HPO22 4 8.30 9.24 9.25 9.90 10.60
12.38d The acidic hydrogen is highlighted in red; bpK1; cpK2; dpK3. In other words, the final pH depends on the equilibrium between HA and A2, Ka 5 [A 2][H 1] HA [2-9] [HA] [A 2] [2-10] so that [H 1] 5 K a We can express [H1] as a pH, and K a as a pK, which yields 2log[H 1] 5 2log K a 2 log [HA] [A 2] [2-11] or pH 5 pK 1 log [A 2] [HA] [2-12]
Equation 2-12 is known as the Henderson–Hasselbalch equation. It relates the pH of a solution to the pK of an acid and the concentration of the acid (HA) and its conjugate base (A2). This equation makes it possible to perform practical calculations to predict the pH of a solution (see Sample Calculation 2-2) or the concentrations of an acid and its
conjugate base at a given pH (see Sample Calculation 2-3). SAMPLE CALCULATION 2-2 Calculate the pH of a 1-L solution to which has been added 6.0 mL of 1.5 M acetic acid and 5.0 mL of 0.4 M sodium acetate. PROBLEM First, calculate the final concentrations of acetic acid (HA) and acetate (A2). The final volume of the solution is 1 L 1 6 mL 1 5
mL 5 1.011 L. SOLUTION (0.006 L)(1.5 M) 5 0.0089 M 1.011 L (0.005 L)(0.4 M) 5 0.0020 M [A 2] 5 1.011 L [HA] 5 Next, substitute these values into the Henderson–Hasselbalch equation using the pK for acetic acid given in Table 2-4: [A 2] [HA] 0.0020 pH 5 4.76 1 log 0.0089 5 4.76 2 0.65 5 4.11 pH 5 pK 1 log 3. Calculate the pH of a 1-L solution to
which has been added 25 mL of 10 mM acetic acid and 25 mL of 30 mM sodium acetate. 4. Calculate the pH of a 500-mL solution to which has been added 10 mL of 50 mM boric acid and 20 mL of 20 mM sodium borate. PRACTICE PROBLEMS SAMPLE CALCULATION 2-3 Calculate the concentration of formate in a 10-mM solution of formic acid at
pH 4.15. PROBLEM The solution of formic acid contains both the acid species (formic acid) and its conjugate base (formate). Use the Henderson–Hasselbalch equation to determine the ratio of formate (A2) to formic acid (HA) at pH 4.15, using the pK value given in Table 2-4. SOLUTION pH 5 pK 1 log [A 2] [HA] [A 2] 5 pH 2 pK 5 4.15 2 3.75 5 0.40
[HA] [A 2] 5 2.51 or [A 2] 5 2.51[HA] [HA] Since the total concentration of formate and formic acid is 0.01 M, [A2] 1 [HA] 5 0.01 M, and [HA] 5 0.01 M 2 [A2]. Therefore, log [A 2] 5 [A 2] 5 [A 2] 5 3.51[A 2] 5 [A 2] 5 2.51[HA] 2.51(0.01 M 2 [A 2]) 0.0251 M 2 2.51[A 2] 0.0251 M 0.0072 M or 7.2 mM 5. Calculate the concentration of acetate in a 50-
mM solution of acetic acid at pH 5.0. 6. Calculate the concentration of phosphoric acid in a 50-mM solution of phosphate at pH 3.0. PRACTICE PROBLEMS The Henderson–Hasselbalch equation indicates that when the pH of a solution of acid is equal to the pK of that acid, then the acid is half dissociated; that is, exactly half of the molecules are in the
protonated HA form and half are in the unprotonated A2 form. You can prove to yourself that when [A2] 5 [HA], the log term of the Henderson–Hasselbalch equation becomes zero (log 1 5 0), and pH 5 pK. When the pH is far below the pK, the acid exists mostly in the HA form; when the pH is far above the pK, the acid exists mostly in the A2 form.
Note that A2 signifies a deprotonated acid; if the acid (HA) bears a positive charge to begin with, the dissociation of a proton yields a neutral species, still designated A2. Knowing the ionization state of an acidic substance at a given pH can be critical. For example, a drug that has no net charge at pH 7.4 may readily enter cells, whereas a drug that
bears a net positive or negative charge at that pH may remain in the bloodstream and be therapeutically useless (see Sample Calculation 2-4). 40 Ch 2 Aqueous Chemistry SAMPLE CALCULATION 2-4 PROBLEM Determine which molecular species of phosphoric acid predominates at pH values of (a) 1.5, (b) 4, (c) 9, and (d) 13. SOLUTION From the pK
values in Table 2-4, we know that: Below pH 2.15, the fully protonated H3PO4 species predominates. At pH 2.15, [H3PO4] 5 [H2PO42]. Between pH 2.15 and 6.82, the H2PO42 species predominates. At pH 6.82, [H2PO42] 5 [HPO22 4]. Between pH 6.82 and 12.38, the HPO22 4 species predominates. 22 32 At pH 12.38, [HPO4] 5 [PO4]. Above pH
12.38, the fully deprotonated PO32 4 species predominates. Therefore, the predominant species at the indicated pH values are (a) H3PO4, 32 (b) H2PO42 , (c) HPO22 4 , and (d) PO4 . PRACTICE PROBLEMS 7. 8. 9. 10. Which species of phosphoric acid predominates at pH 6? Which species of phosphoric acid predominates at pH 8? At what pH will
[NH3] 5 [NH14]? Over what pH range will the HOOCCH2CH2COO2 form of succinic acid predominate? The pH-dependent ionization of biological molecules is also key to understanding their structures and functions. Many of the functional groups on biological molecules act as acids and bases. Their ionization states depend on their respective pK
values and on the pH ([H1]) of their environment. For example, at physiological pH, a polypeptide bears multiple ionic charges because its carboxylic acid (OCOOH) groups are ionized to carboxylate (OCOO2) groups and its amino (ONH2) groups are protonated (ONH31). This is because the pK values for the carboxylic acid groups are about 4, and
the pK values for the amino groups are above 10. Consequently, below pH 4, both the carboxylic acid and amino groups are mostly protonated; above pH 10, both groups are mostly deprotonated. pH , 4 OCOOH ONH13 4 , pH , 10 OCOO2 ONH13 pH . 10 OCOO2 ONH2 Note that a compound containing a OCOO2 group is sometimes still called an
"acid," even though it has already given up its proton. Similarly, a "basic" compound may already have accepted a proton. CONCEPT REVIEW • • • • 41 Tools and Techniques: Buffers What is pH and how is it related to [H1] and [OH2]? How does the addition of an acid or base affect the pH of a solution? What is the relationship between pK and a
molecule's tendency to lose a proton? What is the relationship between pH and the pK of an acid? 2-4 Tools and Techniques: Buffers When a strong acid such as HCl is added to pure water, all the added acid contributes directly to a decrease in pH. But when HCl is added to a solution containing a weak acid in equilibrium with its conjugate base (A2),
the pH does not change so dramatically, because some of the added protons combine with the conjugate base to re-form the acid and therefore do not contribute to an increase in [H1]. KEY CONCEPT • Buffer solutions resist changes in pH. HCl S H 1 1 Cl 2 large increase in [H 1] HCl 1 A 2 n HA 1 Cl 2 small increase in [H 1] Conversely, when a
strong base (such as NaOH) is added to the solution of weak acid/conjugate base, some of the added hydroxide ions accept protons from the acid to form H2O and therefore do not contribute to a decrease in [H1]. NaOH S Na 1 1 OH 2 large decrease in [H 1] NaOH 1 HA S Na 1 1 A 2 1 H2O small decrease in [H 1] The weak acid/conjugate base
system (HA/A2) acts as a buffer against the added acid or base by preventing the dramatic changes in pH that would otherwise occur. The buffering activity of a weak acid, such as acetic acid, can be traced by titrating the acid with a strong base (Fig. 2-17). At the start of the titration, all the acid is present in its protonated (HA) form. As base (for
example, NaOH) is added, protons begin to dissociate from the acid, producing A2. The continued addition of base eventually causes all the protons to dissociate, leaving all the acid in its conjugate base (A2) form. At the midpoint of the titration, exactly half the protons have dissociated, so [HA] 5 [A2] and pH 5 pK (Equation 2-12). The broad, flat
shape of the titration curve shown in Figure 2-17 indicates that the pH does not change drastically with added acid or base when the pH is near the pK. The effective buffering capacity of an acid is generally taken to be within one pH unit of its pK. For acetic acid (pK 5 4.76), this would be pH 3.76–5.76. Figure 2-17 Titration of acetic acid. At the start
point (before base is 8 End point 7 Midpoint [CH3COOH] = [CH3COO−] 6 pK Effective buffering range 5 pH 4 [CH3COOH] > [CH3COO−] 3 2 [CH3COOH] < [CH3COO−] Start point added), the acid is present mainly in its CH3COOH form. As small amounts of base are added, protons dissociate until, at the midpoint of the titration (where pH 5 pK),
[CH3COOH] 5 [CH3COO2]. The addition of more base causes more protons to dissociate until nearly all the acid is in the CH3COO2 form (the end point). The shaded area indicates the effective buffering range of acetic acid. Within one pH unit of the pK, additions of acid or base do not greatly perturb the pH of the solution. See Animated Figure.
Titration curves for acetic acid, phosphate, and ammonia. 1 ? 0 0.5 H+ ions dissociated 1 Sketch the titration curve for ammonia. Biochemists nearly always perform experiments in buffered solutions in order to maintain a constant pH when acidic or basic substances are added or when chemical reactions produce or consume protons. Without
buffering, fluctuations in pH would alter the ionization state of the molecules under study, which might then behave differently. Before biochemists appreciated the importance of pH, experimental results were often poorly reproducible, even within the same laboratory. A buffer solution is typically prepared from a weak acid and the salt of its
conjugate base (see Sample Calculation 2-5). The two are mixed together in the appropriate ratio, according to the Henderson–Hasselbalch equation, and the final pH is adjusted if necessary by adding a small amount of concentrated HCl or NaOH. In addition to choosing a buffering compound with a pK value near the desired pH, a biochemist must
consider other factors, including the compound's solubility, stability, toxicity to cells, reactivity with other molecules, and cost. 42 Ch 2 Aqueous Chemistry SAMPLE CALCULATION 2-5 PROBLEM How many mL of a 2.0-M solution of boric acid must be added to 600 mL of a solution of 10 mM sodium borate in order for the pH to be 9.45? SOLUTION
Rearrange the Henderson–Hasselbalch equation to isolate the [A2]/[HA] term: pH 5 pK 1 log log [A 2] [HA] [A 2] 5 pH 2 pK [HA] [A 2] 5 10(pH2pK) [HA] Substitute the known pK (from Table 2-4) and the desired pH: [A 2] 5 10(9.4529.24) 5 100.21 5 1.62 [HA] The starting solution contains (0.6 L)(0.01 mol ? L21) 5 0.006 moles of borate (A2). The
amount of boric acid (HA) needed is 0.006 mol/1.62 5 0.0037 mol. Since the stock boric acid is 2.0 M, the volume of boric acid to be added is (0.0037 mol)/(2.0 mol ? L21) 5 0.0019 L or 1.9 mL. PRACTICE PROBLEMS 11. How many mL of a 5.0-M solution of sodium borate must be added to a 200-mL solution of 50 mM boric acid in order for the pH to
be 9.6? 12. How many mL of a 1-M solution of imidazolium chloride must be added to a 500-mL solution of 10 mM imidazole in order for the pH to be 6.5? One commonly used laboratory buffer system that mimics physiological conditions contains a mixture of dissolved NaH2PO4 and Na2HPO4 for a total phosphate concentration of 10 mM. The Na1
ions are spectator ions and are usually not significant because the buffer solution usually also contains about 150 mM NaCl (see Fig. 2-13). In this "phosphate-buffered saline," the equilibrium between the two species of phosphate ions can "soak up" added acid (producing more H2PO42) or added base (producing more HPO22 4). pK = 6.82 H2PO 4
H HPO42 Acid added H2PO 4 H HPO42 Base added H2PO 4 H HPO42 This phenomenon illustrates Le Châtelier's principle, which states that a change in concentration of one reactant will shift the concentrations of other reactants in order to restore equilibrium. In the human body, the major buffering systems involve bicarbonate, phosphate, and
other ions (Box 2-D). 43 Tools and Techniques: Buffers CONCEPT REVIEW • How does a solution of a weak acid and its conjugate base minimize changes in pH when a strong acid or base is added to the solution? • What is the useful pH range of a buffer? BOX 2-D CLINICAL CONNECTION Acid–Base Balance in Humans The cells of the human body
typically maintain an internal pH of 6.9–7.4. The body does not normally have to defend itself against strong inorganic acids, but many metabolic processes generate acids, which must be buffered so that they do not cause the pH of blood to drop below its normal value of 7.4. The functional groups of proteins and phosphate groups can serve as
biological buffers; however, the most important buffering system involves CO2 (itself a product of metabolism) in the blood plasma (plasma is the fluid component of blood). CO2 reacts with water to form carbonic acid, H2CO3 acids is initially filtered out of the bloodstream in the kidneys, but the kidneys actively reclaim this bicarbonate before it is
lost in the urine. Most bicarbonate reabsorption is accomplished by a system in which H1 leaves the kidney cells in exchange for Na1 (step 1 in the figure below). The expelled H1 combines with HCO32 in the filtrate, forming CO2 (step 2). Because it is nonpolar, the CO2 can diffuse into the kidney cell, where it is converted back to H 1 1 HCO32 (step
3). Ion exchanger Na⫹ CO2 1 H2O Δ H2CO3 This freely reversible reaction is accelerated in vivo by the enzyme carbonic anhydrase, which is present in most tissues and is particularly abundant in red blood cells. Carbonic acid ionizes to bicarbonate, HCO2 3 (see Box 2-C): HCO⫺ 3 1 3 CO2 so that the overall reaction is 1 1 HCO⫺ 3 H⫹ 2 H2CO3 Δ H
1 1 HCO32 CO2 1 H2O Δ H H⫹ Na⫹ CO2 HCO32 The pK for this process is 6.1 (the ionization of HCO32 to CO22 occurs with a pK of 10.3 and is therefore not significant 3 at physiological pH). Although a pK of 6.1 appears to be just outside the range of a useful physiological buffer (which would be within one pH unit of 7.4), the effectiveness of the
bicarbonate buffer system is augmented by the fact that excess hydrogen ions can not only be buffered but can also be eliminated from the body. This is possible because after the H1 combines with HCO32 to re-form H2CO3, which rapidly equilibrates with CO2 1 H2O, some of the CO2 can be given off as a gas in the lungs. If it becomes necessary to
retain more H1 to maintain a constant pH, breathing can be adjusted so that less gaseous CO2 is lost during exhalation. Changes in pulmonary function can adequately adjust blood pH on the order of minutes to hours; H⫹ pump however, longer-term adjustments of hours to days are made by the kidneys, which use a variety of mechanisms to excrete
or retain H1, bicarbonate, and other ions. In fact, the kidneys play a major role H⫹ in the buffering of metabolic acids. Normal metabolic activity generates acids as the result of the degradation of amino acids, the incomplete oxidation of glucose and fatty acids, and the ingestion of acidic groups in the form of phosphoproteins and phosFILTRATE
pholipids. The HCO2 3 required to buffer these FILTRATE KIDNEY CELL In addition to reabsorbing filtered HCO2 3 , the kidneys also generate additional HCO2 to offset losses due to the buffering 3 of metabolic acids and the exhalation of CO2. Metabolic activity in the kidney cells produces CO2, which is converted to H 1 1 HCO32 . The cells
actively secrete the H1, which is lost via the urine (step 1 in the figure below), accounting for the mildly acidic pH of normal urine. The bicarbonate remaining in the cell is returned to the bloodstream in exchange for Cl2 (step 2). Ion exchanger Cl⫺ HCO3⫺ H⫹ 1 Cl⫺ 2 HCO⫺ 3 CO2 KIDNEY CELL BLOOD Additional HCO32 accumulates as a result of
the metabolism of the amino acid glutamine in the kidneys. COO H C O CH2 CH2 C NH2 NH3 Glutamine The two amino groups are removed as ammonia (NH3), which is ultimately excreted in the urine. Because the ammonium ion (NH41) has a pK value of 9.25, nearly all the ammonia molecules become protonated at physiological pH. The
consumption of protons from carbonic acid (ultimately, CO2) leaves an excess of HCO2 3. Certain medical conditions can disrupt normal acid–base balance, leading to acidosis (blood pH less than 7.35) or alkalosis (blood pH greater than 7.45). The activities of the lungs and kidneys, as well as other organs, may contribute to the imbalance or respond
to help correct the imbalance. The most common disorder of acid–base chemistry is metabolic acidosis, which is caused by the accumulation of the acidic products of metabolism and can develop during shock, starvation, severe diarrhea (in which bicarbonate-rich digestive fluid is lost), certain genetic diseases, and renal failure (in which damaged
kidneys eliminate too little acid). Metabolic acidosis can interfere with cardiac function and oxygen delivery and contribute to central nervous system depression. Despite its many different causes, one common symptom of metabolic acidosis is rapid, deep breathing. The increased ventilation helps compensate for the acidosis by "blowing off " more
acid in the form of CO2 derived from H2CO3. However, this mechanism also impairs O2 uptake by the lungs. Metabolic acidosis can be treated by administering sodium bicarbonate (NaHCO3). In chronic metabolic acidosis, the mineral component of bone serves as a buffer, which leads to the loss of calcium, magnesium, and phosphate and ultimately
to osteoporosis and fractures. Normal conditions H HCO 3 H2CO3 H2O CO2 Excess acid H⫹ HCO 3 H HCO 3 H HCO 3 H2CO3 H2CO3 H2CO3 H2O CO2 H2O CO2 H2O CO2 Insufficient acid H HCO 3 H HCO 3 H⫹ HCO 3 H2CO3 H2CO3 H2CO3 H2O CO2 H2O CO2 H2O CO2 Metabolic alkalosis, a less common condition, can result from
prolonged vomiting, which represents a loss of HCl in gastric fluid. This specific disorder can be treated by infusing NaCl. Metabolic alkalosis is also caused by overproduction of mineralocorticoids (hormones produced by the adrenal glands), which leads to abnormally high levels of H1 excretion and Na1 retention. This disorder does not respond to
saline infusion. Individuals with metabolic alkalosis may experience apnea (cessation of breathing) for 15 s or more and cyanosis (blue coloration) due to inadequate oxygen uptake. Both these symptoms reflect the body's effort to compensate for the high blood pH by minimizing the loss of CO2 from the lungs. What happens when abnormal lung
function is the cause of an acid–base imbalance? Respiratory acidosis can result from impaired pulmonary function caused by airway blockage, asthma (constriction of the airways), and emphysema (loss of alveolar tissue). In all cases, the kidneys respond by adjusting their activity, mainly by increasing the synthesis of the enzymes that break down
glutamine in order to produce NH3. Excretion of NH41 helps correct the acidosis. However, renal compensation of respiratory acidosis takes hours to days (the time course for adjusting enzyme levels), so this condition is best treated by restoring pulmonary function through bronchodilation, supplemental oxygen, or mechanical ventilation (assisted
breathing). Some of the diseases that impair lung function (for example, asthma) can also contribute to respiratory alkalosis, but this relatively rare condition is more often caused by hyperventilation brought on by fear or anxiety. Unlike the other acid–base disorders, this form of respiratory alkalosis is seldom life-threatening. Questions: 1. Impaired
pulmonary function can contribute to respiratory acidosis. Using the appropriate equations, explain how the failure to eliminate sufficient CO2 through the lungs leads to acidosis. 2. Metabolic acidosis often occurs in patients with impaired circulation from cardiac arrest. Mechanical hyperventilation (a standard treatment for acidosis) cannot be used
with these patients because they often have acute lung injury (ALI). A group of physicians at San Francisco General Hospital advocated using tris(hydroxymethyl)aminomethane (Tris) to treat the metabolic acidosis of these patients. (a) How would mechanical hyperventilation help alleviate acidosis in patients who do not have ALI? (b) Why would
treatment with sodium bicarbonate be effective in treating metabolic acidosis? Why would this treatment be unacceptable for patients with ALI? (c) Explain how Tris works to treat metabolic acidosis. Why is this treatment acceptable for patients with ALI? CH2OH HOH2C C NH2 CH2OH Tris(hydroxymethyl)aminomethane [From Kallet, R. H., et al.,
Am. J. Respir. Crit. Care Med. 161, 1149–1153 (2000).] 3. An individual who develops alkalosis by hyperventilating is encouraged to breathe into a paper bag for several minutes. Why does this treatment correct the alkalosis? 4. Ammonia produced in kidney cells can diffuse out of the cell, but only in its unprotonated (NH3) form. (a) Explain. (b) The
NH41 ion, however, can exit the cell in exchange for another ion. What ion is most likely involved? 5. Kidney cells excrete H1 and reabsorb HCO32 from the filtrate. The movement of each of these ions is thermodynamically unfavorable. However, the movement of each becomes possible when it is coupled to another, thermodynamically favorable ion
transport process. Explain how the movement of other ions drives the transport of H1 and HCO32 . 6. In uncontrolled diabetes, the body converts fats to the socalled ketone bodies acetoacetate and 3-hydroxybutyrate, which accumulate in the bloodstream. O CH3 C OH CH2 COO CH3 C CH2 COO H Acetoacetate 3-Hydroxybutyrate Do the ketone
bodies contribute to acidosis or alkalosis? How might the body compensate for this acid–base imbalance? 7. Kidney cells have a carbonic anhydrase on their external surface as well as an intracellular carbonic anhydrase. What are the functions of these two enzymes? 8. Explain why the lungs can rapidly compensate for metabolic acidosis, whereas the
kidneys are slow to compensate for respiratory acidosis. SUMMARY 2-1 Water Molecules Form Hydrogen Bonds 2-3 Acid–Base Chemistry • Water molecules are polar; they form hydrogen bonds with each other and with other polar molecules bearing hydrogen bond donor or acceptor groups. • The electrostatic forces acting on biological molecules
also include ionic interactions and van der Waals interactions. • Water dissolves polar and ionic substances. • The dissociation of water produces hydroxide ions (OH2) and protons (H1) whose concentration can be expressed as a pH value. The pH of a solution can be altered by adding an acid (which donates protons) or a base (which accepts
protons). • The tendency for a proton to dissociate from an acid is expressed as a pK value. • The Henderson–Hasselbalch equation relates the pH of a solution of a weak acid and its conjugate base to the pK and the concentrations of the acid and base. 2-2 The Hydrophobic Effect • Nonpolar (hydrophobic) substances tend to aggregate rather than
disperse in water in order to minimize the decrease in entropy that would be required for water molecules to surround each nonpolar molecule. This is the hydrophobic effect. • Amphiphilic molecules, which contain both polar and nonpolar groups, may aggregate to form micelles or bilayers. 2-4 Tools and Techniques: Buffers • A buffered solution,
which contains an acid and its conjugate base, resists changes in pH when more acid or base is added. GLOSSARY TERMS polarity hydrogen bond ionic interaction van der Waals radius electronegativity van der Waals interaction dipole–dipole interaction London dispersion forces dielectric constant solute solvation hydration hydrophilic hydrophobic
hydrophobic effect amphiphilic amphipathic micelle bilayer vesicle hydronium ion proton jumping ionization constant of water (K w) neutral solution acidic solution basic solution pH acid base acid dissociation constant (K a) pK conjugate base polyprotic acid Henderson–Hasselbalch equation buffer Le Châtelier's principle acidosis alkalosis metabolic
acidosis metabolic alkalosis respiratory acidosis respiratory alkalosis PROBLEMS 2-1 Water Molecules Form Hydrogen Bonds 1. The HOCOH bond angle in the perfectly tetrahedral CH4 molecule is 1098. Explain why the HOOOH bond angle in water is only about 104.58. 2. Each CPO bond in CO2 is polar, yet the whole molecule is nonpolar. Explain.
3. Is ammonia a polar molecule? 4. Consider the following molecules and their melting points listed below. How can you account for the differences in melting points among these molecules of similar size? Water, H2O Ammonia, NH3 Methane, CH4 Molecular weight (g ? mol21) Melting point (8C) 18.0 17.0 16.0 0 277 2182 46 5. Identify the hydrogen
bond acceptor and donor groups in the following molecules. Use an arrow to point toward each acceptor and away from each donor. O (c) H2N C (d) H3C (CH2)3 NH2 CH3 O H3N CH (e) H3C CH2 O O C N CH2 CH C O COO H O O S N N Aspartame H2N H O O N N H H NH2 Uric acid O Sulfanilamide 6. Do intermolecular hydrogen bonds form in
the compounds below? Draw the hydrogen bonds where appropriate. O (a) H3C C H (b) N (c) H3C OH CH2 Cl and H (e) H3C CH2 O CH2 O H CH3 and H N N H 7. What are the most important intermolecular interactions in the following molecules? C CH3 O (b) H3C C (c) H3C CH2 14. The amino acid glycine is sometimes drawn as structure A below.
However, the structure of glycine is more accurately represented by structure B. Glycine has the following properties: white crystalline solid, high melting point, and high water solubility. Why does structure B more accurately represent the structure of glycine than structure A? CH3 8. Examine Table 2-1. (a) Rank the five atoms listed in order of
increasing electronegativity. (b) What is the relationship between an atom's electronegativity and its ability to participate in hydrogen bonding? 9. Rank the melting points of the following compounds: CH2 O O C CH2 COOH NH2 H3N CH2 COO Structure B 15. (a) Water has a surface tension that is nearly three times greater than that of ethanol.
Explain. (b) The surface tension of water has been shown to decrease with increasing temperature. Explain. 16. Explain why water forms nearly spherical droplets on the surface of a freshly waxed car. Why doesn't water bead on a clean windshield? 18. Editors of a popular sporting magazine recommend sponging skin down with isopropyl alcohol
after a noontime bike ride. Why is this a quick alternative to a shower? NH2 (d) CsCl (b) H3C 13. Ammonium sulfate, (NH4)2SO4, is a water-soluble salt. Draw the structures of the hydrated ions that form when ammonium sulfate dissolves in water. 17. Which of the four primary alcohols listed in Table 2-2 would be the best solvent for ammonium
ions? What can you conclude about the polarity of these solvents, which all contain an OOH group that can form hydrogen bonds? O (a) H3C 12. Ice skaters skate not on the solid ice surface itself but on a thin film of water that forms between the skater's blade and the ice. What unique property of water makes ice skating possible? Structure A (d)
H3C (a) H3C 11. Water is unusual in that its solid form is less dense than its liquid form. This means that when a pond freezes in the winter, ice is found as a layer on top of the pond, not on the bottom. What are the biological advantages of this? H2N N H CH2 CH 10. Classify the compounds listed in Problem 9 as soluble, slightly soluble, or insoluble
in water and explain your answer. CH3 H CH2 CH3 19. A typical spherical bacterial cell with a diameter of 1 mm contains 1000 molecules of a certain protein. (a) What is the protein's concentration in units of mM? (Hint: The volume of a sphere is 4pr3/3.) (b) How many molecules of glucose does the cell contain if the internal glucose concentration is
5 mM? 20. Practitioners of homeopathic medicine believe that good health can be restored by administering a small amount of a harmful substance that stimulates the body's natural healing processes. The remedy is typically prepared by serially diluting an animal or plant extract. In a so-called 303 dilution, the active substance is diluted 10-fold, 30
times in succession. (a) If the substance is initially present at a concentration of 1 M, what is its concentration after a 303 dilution? 47 (b) A typical homeopathic dose is a few drops. How many molecules of the active substance would be present in 1 mL? (c) When confronted with the results of an exercise such as the one in part (b), proponents of
homeopathy claim that a molecule can leave an imprint, or memory of itself, on the surrounding water molecules. What information presented in this chapter might be used to support this claim? To refute it? 2-2 The Hydrophobic Effect 21. Consider the structures of the molecules below. Are these molecules polar, nonpolar, or amphiphilic? Which are
capable of forming micelles? Which are capable of forming bilayers? CH3 (a) H3C N (CH2)11 CH2COO CH3 (b) H3C (CH2)11 (c) H3C CH3 Cl N CH3 CH3 O O CH2 HC O C C (CH2)11 (CH2)11 HO CH2 O (e) H3C CH COO CH3 CH3 OH 22. The compound bis-(2-ethylhexyl)sulfosuccinate (abbreviated AOT) is capable of forming "reverse" micelles in
the hydrocarbon solvent isooctane (2,2,4-trimethylpentane). Scientists have investigated the use of reverse micelles for extracting water-soluble proteins. A two-phase system is formed: the hydrocarbon phase containing the reverse micelles and the water phase containing the protein. After a certain period of time, the protein is transferred to the
reverse micelle. (a) Draw the structure of the reverse micelle that AOT would form in isooctane. (b) Where would the protein be located in the reverse micelle? CH2CH3 H 3C H3C (CH2)3 (CH2)3 CH CH CH2 CH2 CH2CH3 H 3C O C C O S ONa O Sodium dodecyl sulfate (SDS) 24. A device sometimes called a laundry ball is advertised as a
replacement for environmentally harmful detergent. The laundry ball is a rubber-coated, baseball-sized plastic sphere containing fluid that is said to serve as an organizing template for water molecules. (a) Evaluate this claim based on your knowledge of water structure. (b) Magnets inside the laundry ball are said to rip apart clusters of water
molecules to make it easier for individual water molecules to access and remove dirt molecules. If true, would this action help wash away dirt? (c) The instructions recommend that a laundry ball, rather than the usual detergent, be added to a load of laundry. Why do the instructions recommend washing with hot rather than cold water? 26. Fresh

water is obtained from seawater in a desalination process using reverse osmosis. In reverse osmosis, seawater is forced through a membrane; the salt remains on one side of the membrane and fresh water is produced on the other side. In what ways does reverse osmosis differ from osmosis as described in Problem 25? 27. Which of the following substances might be able to cross a bilayer? Which substances could not? Explain your answer. (a) CO2 (b) CHO H HO OH H H OH H OH CH2OH O O (CH2)11 25. Just as a dissolved substance tends to move spontaneously down its concentration gradient, water also tends to move from an area of high concentration (low solute concentration) to an area of low concentration (high solute concentration), a process known as osmosis. (a) Explain why a lipid bilayer would be a barrier to osmosis. (b) Why are isolated human cells placed in a solution that typically contains about 150 mM NaCl? What would happen if the cells were placed in pure water? CH3 (d) O Glucose CH2 O CH O S O O Bis-(2-ethylhexyl)sulfosuccinate (AOT) 23. Many household soaps are amphiphilic substances, often the salts of long-chain fatty acids, that form water-soluble micelles. An example is sodium dodecyl sulfate (SDS above, right). (a) Identify the polar and nonpolar portions of the SDS molecule. (b) Draw the structure of the micelle formed by SDS. (c) Explain how the SDS micelles "wash away" water-insoluble substances such as cooking grease. (c) OH NO2 NO2 2,4-Dinitrophenol (DNP) (d) Ca21 28. Drug delivery is a challenging problem because the drug molecules must be sufficiently water soluble to dissolve in the blood, but also sufficiently nonpolar to pass through the cell membrane for delivery. A medicinal chemist proposes to encapsulate a watersoluble drug into a vesicle (see Section 2-2). How does this strategy facilitate the delivery of the drug to its target cell? 29. A specialized protein pump in the red blood cell membrane exports Na1 ions and imports K1 ions in order to maintain the sodium and potassium ion concentrations shown in Figure 2-13. Does the movement of these ions occur spontaneously, or is this an energy-requiring process? Explain. 48 30. Estimate the amount of Na1 lost in sweat during 15 minutes of vigorous exercise. What is the mass of potato chips (200 mg Na1 per ounce) you would have to consume in order to replace the lost sodium? 31. The bacterium E. coli has the ability to change both its water content and its cytoplasmic K1 concentration to adapt to changes in the solute concentration of its growth medium. (The solute concentration is referred to as osmolarity—a solution with a high concentration of solute has a high osmolarity and a solution with a low concentration of solute has a low osmolarity.) The bacterial cell consists of a cytoplasmic compartment bounded by a membrane, which is in turn surrounded by a cell wall. The cell wall is porous and can accommodate the passage of water and ions. (The membrane is not as porous but does allow passage of water and ions.) In addition, the cell wall is also fairly elastic and can stretch to accommodate increases in the cytoplasmic compartment volume. E. coli Periplasmic space Cytoplasm Cytoplasmic membrane 41. Calculate the pH of 500 mL of water to which (a) 20 mL of 1.0 M HNO3 or (b) 15 mL of 1.0 M KOH has been added. 42. What is the pH of 1.0 L of water to which (a) 1.5 mL of 3.0 M HCl or (b) 1.5 mL of 3.0 M NaOH has been added? 43. Several hours after a meal, partially digested food leaves the stomach and enters the small intestine, where pancreatic juice is added. How does the pH of the partially digested mixture change as it passes from the stomach into the intestine (see Table 2-3)? 44. The pH of urine has been found to be correlated with diet. Acidic urine results when meat and dairy products are consumed, because the oxidation of sulfur-containing amino acids in the proteins produces protons. Consumption of fruits and vegetables leads to alkaline urine, because these foods contain plentiful amounts of potassium and magnesium carbonate salts. Why would the presence of these salts result in alkaline urine? 45. Give the conjugate base of the following acids: (e) HAsO22 (a) HC2O42 4 2 (b) HSO3 (f) HPO22 4 (c) H2PO42 (g) HO22 2 (d) HCO3 46. Give the conjugate acid of the species listed in Problem 45. 47. Identify the acidic hydrogens in the following compounds. Cell wall Under nongrowing conditions, E. coli responds to changes in osmolarity by regulating its cytoplasmic water content. What would happen to the cytoplasmic volume if E. coli were grown in a highosmolarity medium? In a low-osmolarity medium? 32. Under growth conditions, E. coli also regulates cytoplasmic K1 content (in addition to water) in response to growth medium osmolarity (see Problem 31). This prevents the large changes in cell volume that would occur in response to changes in growth medium osmolarity if only water content were regulated. How might E. coli regulate both the cytoplasmic concentrations of K1 and water when grown in a low-osmolarity medium? What happens when E. coli is grown in a high-osmolarity medium? CH2 HO C Piperidine O H3N CH C OH HO O O C C CH2 Oxalic acid CH2 O CH2 35. What is the pH of a solution of 1.0 3 1029 M HCl? 36. What is the pH of a solution of 1.0 3 1029 M NaOH? 37. Explain why H3O1 is the strongest acid that can exist in a biological system. 38. Draw a diagram similar to Fig. 2-14 showing how a hydroxide ion appears to jump through an aqueous solution. 39. Calculate [H1] for saliva (pH ≈6.6) and urine (pH ≈5.5). 40. Fill in the blanks of the following table: [H1] (M) [OH2] (M) 4.5 3 1027 O NH3 34. Like all equilibrium constants, the value of K w is temperature dependent. Kw is 1.47 3 10214 at 308C. What is "neutral" pH at this temperature? Lysine 2.1 3 10 23 O N H Barbituric acid O O NH CH2 CH2 S O O 4-Morphine ethanesulfonic acid (MES) 48. The structure of pyruvic acid is shown below. (a) Draw the structure of pyruvate. (b) Using what you have learned about acidic functional groups, which form of this compound is likely to predominate in the cell at pH 7.4? Explain. CH3 C O COOH Neutral OH NH CH2 33. Compare the concentrations of H2O and H in a sample of pure water at pH 7.0 at 258C. A B C D N H2 COOH Citric acid 1 pH 5.60 COOH CH2 2-3 Acid–Base Chemistry Acid, base, or neutral? COOH Pyruvic acid 49 49. The structure of a neutral form of lysine is shown below. Use the information presented in Section 2-3 to draw the structure of this compound at pH 7. O O H2N CH C C OH C CH3 O CH2 Acetylsalicylic acid (aspirin) CH2 59. A solution is made by mixing 50 mL of a stock solution of 2.0 M K2HPO4 and 25 mL of a stock solution of 2.0 M KH2PO4. The solution is diluted to a final volume of 200 mL. What is the final pH of the solution? CH2 NH2 50. Rank the following according to their strength as acids: A B C D E pK a = 2.97 OH O CH2 Acid Citrate Succinic acid Succinate Formic acid Citric acid 58. The structure of acetylsalicylic acid (aspirin) is shown. Is aspirin more likely to be absorbed (pass through a lipid membrane) in the stomach (pH ≈2) or in the small intestine (pH ≈8)? Explain. Ka pK 4.76 6.17 3 1025 2.29 3 1026 1.78 3 1024 3.13 51. The pK of CH3CH2NH1 3 is 10.7. Would the pK of FCH2CH2NH1 3 be higher or lower? 52. The amino acid glycine (H2NOCH2OCOOH) has pK values of 2.35 and 9.78. Indicate the structure and net charge of the molecular species that predominates at pH 2, 7, and 10. 2-4 Tools and Techniques: Buffers 53. Which would be a more effective buffer at pH 8.0 (see Table 2-4)? (a) 10 mM HEPES buffer or 10 mM glycinamide buffer (b) 10 mM Tris buffer or 20 mM Tris buffer (c) 10 mM boric acid or 10 mM sodium borate 54. Which would be a more effective buffer at pH 5.0 (see Table 2-4)? (a) 10 mM acetic acid buffer or 10 mM HEPES buffer (b) 10 mM acetic acid buffer or 20 mM acetic acid buffer (c) 10 mM acetic acid or 10 mM sodium acetate 55. Explain why metabolically generated CO2 does not accumulate in tissues but is quickly converted to carbonic acid by the action of carbonic anhydrase in red blood cells. 56. The pH of blood is maintained within a narrow range (7.35–7.45). Carbonic acid, H2CO3, participates in blood buffering. (a) Write the equations for the dissociation of the two ionizable protons. (b) The pK for the first ionizable proton is 6.35; the pK for the second ionizable proton is 10.33. Use this information to identify the weak acid and the conjugate base present in the blood. 57. Phosphoric acid, H3PO4, has three ionizable protons. (a) Sketch the titration curve. Indicate the pK values and the species that predominate in each area of the curve. (b) Which two phosphate species are present in the blood at pH 7.4? (c) Which two phosphate species would be used to prepare a buffer solution of pH 11? 60. Calculate the ratio of imidazole to the imidazolium ion in a solution at pH 7.4. 61. What is the volume (in mL) of glacial acetic acid (17.4 M) that would have to be added to 500 mL of a solution of 0.20 M sodium acetate in order to achieve a pH of 5.0? 62. What is the mass of NaOH that would have to be added to 500 mL of a solution of 0.20 M acetic acid in order to achieve a pH of 5.0? 63. An experiment requires the buffer HEPES, pH 5 8.0 (see Table 2-4). (a) Write an equation for the dissociation of HEPES in water. Identify the weak acid and the conjugate base. (b) What is the effective buffering range for HEPES? (c) The buffer will be prepared by making 1.0 L of a 0.10 M solution of HEPES. Hydrochloric acid will be added until the desired pH is achieved. Describe how you will make 1.0 L of 0.10 M HEPES. (HEPES is supplied by a chemical company as a sodium salt with a molecular weight of 260.3 g ? mol21.) (d) What is the volume (in mL) of a stock solution of 6.0 M HCl that must be added to the 0.1 M HEPES to achieve the desired pH of 8.0? Describe how you will make the buffer. 64. One liter of a 0.10 M Tris buffer (see Table 2-4) is prepared and adjusted to a pH of 8.2. (a) Write the equation for the dissociation of Tris in water. Identify the weak acid and the conjugate base. (b) What is the effective buffering range for Tris? (c) What are the concentrations of the conjugate acid and weak base at pH 8.2? (d) What is the ratio of conjugate base to weak acid if 1.5 mL of 3.0 M HCl is added to 1.0 L of the buffer? What is the new pH? Has the buffer functioned effectively? Compare the pH change to that of Problem 42a in which the same amount of acid was added to the same volume of pure water. (e) What is the ratio of conjugate base to weak acid if 1.5 mL of 3.0 M NaOH is added to 1.0 L of the buffer? What is the new pH? Has the buffer functioned effectively? Compare the pH change to that of Problem 42b in which the same amount of base was added to the same volume of pure water. 65. One liter of a 0.1 M Tris buffer (see Table 2-4) is prepared and adjusted to a pH of 2.0. (a) What are the concentrations of the conjugate base and weak acid at this pH? (b) What is the pH when 1.5 mL of 3.0 M HCl is added to 1.0 L of the buffer? Has the buffer functioned effectively? Explain. 50 (c) What is the pH when 1.5 mL of 3.0 M NaOH is added to 1.0 L of the buffer? Has the buffer functioned effectively? Explain. 66. A patient who has taken an overdose of aspirin is brought into the emergency room for treatment. She suffers from respiratory alkalosis, and the pH of her blood is 7.5. Determine the ratio of HCO32 to H2CO3 in the patient's blood at this pH. How does this compare to the ratio of HCO32 to H2CO3 in normal blood? Can the H2CO3/HCO32 system work effectively as a buffer in this patient under these conditions? blood pH from 7.4 to 7.35 or below. The kidney plays a vital role in regulating blood pH. The kidney can either excrete or reabsorb various ions, including phosphate, H2PO42; ammonium, NH1 4 ; or bicarbonate, HCO2 3 . Which ions are excreted and which ions are reabsorbed in metabolic acidosis? Explain, using relevant chemical equations. 68. Metabolic alkalosis occurs when the blood pH rises to 7.45 or greater. Which ions are excreted and which ions are reabsorbed in metabolic alkalosis (see Problem 67)? 67. Metabolic acidosis is a general term that describes a number of metabolic disorders in the body that result in a lowering of the

SELECTED READINGS Ellis, R. J., Macromolecular crowding: obvious but underappreciated, Trends Biochem. Sci. 26, 597–603 (2001). [Discusses how the large number of macromolecules in cells could affect reaction equilibria and reaction rates.] Gerstein, M., and Levitt, M., Simulating water and the molecules of life, Sci. Am. 279(11), 101–105 (1998). [Describes the structure of water and how water interacts with other molecules.] Halperin, M. L., and Goldstein, M. B., Fluid, Electrolyte, and Acid–Base Physiology: A Problem-Based Approach (3rd ed.), W. B. Saunders (1999). [Includes extensive problem sets with explanations of basic science as well as clinical effects of acid–base disorders.] Kropman, M. F., and Bakker, H. J., Dynamics of water molecules, Science 291, 2118–2120 (2001). [Describes how water molecules in solvation shells move more slowly than bulk water molecules.] Yucha, C., Renal regulation of acid–base balance, Nephrol. Nursing J. 31, 201–206 (2004). [A short review of physiological buffer systems and the role of the kidney.]

chapter 3 FROM GENES TO PROTEINS HOW do researchers decipher the information in DNA? Virtually all biological materials, such as this blood sample, contain DNA. Even minute amounts of biological fluids can yield traces of DNA. Because the sequence of nucleotides in a DNA molecule is a form of information, the ability to read and interpret the sequence makes it possible to identify an individual or diagnose a disease. For example, analysis of fetal DNA in the maternal circulation can reveal chromosomal abnormalities as well as the presence of sequence variants that can cause disease. Yet the techniques for reading DNA sequences, described in this chapter, also generate challenges in compiling vast quantities of data and understanding how cells themselves store and use that information. [Lewis Wright/iStockphoto.] THIS CHAPTER IN CONTEXT Do You Remember? Part 1 Foundations Part 2 Molecular Structure and Function 3 From Genes to Proteins Cells contain four major types of biological molecules and three major types of polymers (Section 1-2). Modern prokaryotic and eukaryotic cells apparently evolved from simpler nonliving systems (Section 1-4). Noncovalent forces, including hydrogen bonds, ionic interactions, and van der Waals forces, act on biological molecules (Section 2-1). Part 3 Metabolism Part 4 Genetic Information 51 52 Ch 3 From Genes to Proteins All the structural components of cells and the machinery that carries out the cell's activities are ultimately specified by the cell's genetic material—DNA. Therefore, before examining other types of biological molecules and their metabolic transformations, we must consider the nature of DNA, including its chemical structure and how its biological information is organized and expressed. The Tools and Techniques section of this chapter includes some of the methods used to study and manipulate DNA in the laboratory. 3-1 DNA Is the Genetic Material KEY CONCEPTS • DNA and RNA are polymers of nucleotides, each of which consists of a purine or pyrimidine base, deoxyribose or ribose, and phosphate. • A DNA molecule contains two antiparallel strands that wind around each other to form a double helix in which A and T bases in opposite strands, and C and G bases in opposite strands, pair through hydrogen bonding. • Double-stranded nucleic acids are denatured at high temperatures; at lower temperatures, complementary polynucleotides anneal. Gregor Mendel was certainly not the first to notice that an organism's characteristics (for example, flower color or seed shape in pea plants) were passed to its progeny, but in 1865 he was the first to describe their predictable patterns of inheritance. By 1903, Mendel's inherited factors (now called genes) were recognized as belonging to chromosomes (a word that means "colored bodies"), which are visible by light microscopy (Fig. 3-1). Nucleic acids had been discovered in 1869 by Friedrich Miescher, who isolated this material from the white blood cells in pus on surgical bandages. However, when it became clear that chromosomes were composed of both proteins and nucleic acids, nucleic acids were dismissed as possible carriers of genetic information due to their lack of complexity: They contained only four different types of structural units, called nucleotides. In contrast, proteins contained 20 different types of amino acids and exhibited great diversity in composition, size, and shape—attributes that seemed more appropriate for carriers of genetic information. Years later, microbiologists showed that a substance from a dead pathogenic (disease-causing) strain of the bacterium Streptococcus pneumoniae could transform cells from a normal strain to the pathogenic type. In 1944, Oswald Avery, Colin MacLeod, and Maclyn McCarty showed that this transforming substance was deoxyribonucleic acid (DNA), but their results did not garner much attention. Another seven years went by until Alfred Hershey and Martha Chase demonstrated that in bacteriophages (viruses that infect bacterial cells and that consist only of protein and DNA), the DNA, not the protein, was the infectious agent (Fig. 3-2). By this time, DNA was known to contain chains of polymerized nucleotides— abbreviated A, C, G, and T—but these were thought to occur as simple repeating tetranucleotides, for example, —ACGT-ACGT-ACGT-ACGT-ACGT— Figure 3-1 Human chromosomes from amniocentesis. In this image, the chromosomes have been stained with fluorescent dyes. [Dr. P. Boyer/Photo Researchers, Inc.] Figure 3-2 A T-type bacteriophage. The phage consists mostly of a protein coat surrounding a molecule of DNA. Alfred Hershey and Martha Chase identified the DNA as the infectious agent. [Dept. of Microbiology, Biozentrum/Science Photo Library/Photo Researchers.] When Erwin Chargaff showed in 1950 that the nucleotides in DNA were not all present in equal numbers and that the nucleotide composition varied among species, it became apparent that DNA might be complex enough to be the genetic material after all, and the race was on to decipher its molecular structure. The DNA structure ultimately elucidated by James Watson and Francis Crick in 1953 incorporated Chargaff 's observations. Specifically, Chargaff noted that the amount of A is equal to the amount of T, the amount of C is equal to the amount of G, and the total amount of A ⩽ G is equal to the total amount of C ⩽ T. Chargaff 's "rules" could be satisfied by a molecule with two polynucleotide strands (polymers of nucleotides) in which A and C in one strand pair with T and G in the other. Nucleic acids are polymers of nucleotides Each nucleotide of DNA includes a nitrogen-containing base. The bases adenine (A) and guanine (G) are known as purines because they resemble the organic compound purine: O NH2 H N N N N N1 2 N N H2N H 7 5 4 3 N N N 6 9 8 N N H H Adenine Guanine Purine The bases cytosine (C) and thymine (T) are known as pyrimidines because they resemble the organic compound pyrimidine: O NH2 H N CH3 N 4 N3 2 O O N H N 5 6 1 N H Cytosine Thymine Pyrimidine Ribonucleic acid (RNA) contains the pyrimidine uracil (U) rather than thymine: O H N O N H Uracil so that DNA contains the bases A, C, G, and T, whereas RNA contains A, C, G, and U. Linking atom N9 in a purine or atom N1 in a pyrimidine to a five-carbon sugar forms a nucleoside. In DNA, the sugar is 29-deoxyribose; in RNA, the sugar is ribose (the sugar atoms are numbered with primes to distinguish them from the atoms of the attached bases). 5 Ribose 4 H H 3 OH 5 base HOCH2 O H 1 2 OH H base HOCH2 O 2-Deoxyribose 4 H H 3 OH H 1 2 H H 53 DNA Is the Genetic Material 54 Ch 3 From Genes to Proteins TABLE 3-1 Nucleic Acid Bases, Nucleosides, and Nucleotides Base Nucleosidea Nucleotidesa Adenine (A) Adenosine Cytosine (C) Cytidine Guanine (G) Guanosine Thymine (T)b Thymidine Uracil (U)c Uridine Adenylate; adenosine monophosphate (AMP) adenosine diphosphate (ADP) adenosine triphosphate (ATP) Cytidylate; cytidine monophosphate (CMP) cytidine diphosphate (CDP) cytidine triphosphate (CTP) Guanylate; guanosine monophosphate (GMP) guanosine diphosphate (GDP) guanosine triphosphate (GTP) Thymidylate; thymidine monophosphate (TMP) thymidine diphosphate (TDP) thymidine triphosphate (TTP) Uridylate; uridine monophosphate (UMP) uridine diphosphate (UDP) uridine triphosphate (UTP) a Nucleosides and nucleotides containing 29-deoxyribose rather than ribose may be called deoxynucleosides and deoxynucleotides. The nucleotide abbreviation is then preceded by "d." b Thymine is found in DNA but not in RNA. c Uracil is found in RNA but not in DNA. A nucleotide is a nucleoside to which one or more phosphate groups are linked, usually at C5′ of the sugar. Depending on whether there are one, two, or three phosphate groups, the nucleotide is known as a nucleoside monophosphate, nucleoside diphosphate, or nucleoside triphosphate and is represented by a three-letter abbreviation, for example, H H H H OH OH Adenosine monophosphate (AMP) N O P O P O H2C O O O O H H N N N H 2N O O O P O H2C O N N O H N N NH2 O NH2 N O O O O H H H N H H OH OH OH OH Guanosine diphosphate (GDP) O O O P O P O P O H2C O H O Cytidine triphosphate (CTP) Deoxynucleotides are named in a similar fashion, and their abbreviations are pre- ceded by "d." The deoxy counterparts of the compounds shown above would therefore be deoxyadenosine monophosphate (dAMP), deoxyguanosine diphosphate (dGDP), and deoxycytidine triphosphate (dCTP). The names and abbreviations of the common bases, nucleosides, and nucleotides are summarized in Table 3-1. Some nucleotides have other functions In addition to serving as the building blocks for DNA and RNA, nucleotides perform a variety of functions in the cell. They are involved in energy transduction, intracellular signaling, and regulation of enzyme activity. Some nucleotide derivatives are essential players in the metabolic pathways that synthesize biomolecules or degrade them in order to "capture" free energy. For example, coenzyme A (CoA; Fig. 3-3a) is a carrier of other molecules during their synthesis and degradation. Two nucleotides are linked in the compounds nicotinamide adenine dinucleotide (NAD; Fig. 3-3b) and flavin adenine dinucleotide (FAD; Fig. 3-3c), which undergo reversible oxidation and reduction during a number of metabolic reactions. Interestingly, a portion of the structures of each of these molecules is derived from a vitamin, a compound that must be obtained from the diet. O HN C CH2 CH2 C Nicotinamide SH O N CH2 CH2 O O CH2 Ribose H NH Pantothenic acid residue HO C O C H H3C C NH2 NH2 O CH2 O P O O P O O CH2 H O N Coenzyme A (CoA) O N O O H O OH O P O H CH2 H O Niacin N N N O H Adenosine H HO O N OH N O Adenosine H NH2 O P N H P H HO N CH3 O H COO H OH Nicotinamide adenine dinucleotide (NAD) (a) (b) NH2 N N O O CH2 HO Riboflavin C O O H HO C H HO C H O P P OCH2 O N N O H Adenosine H H H OH OH CH2 H3C N H 3C N N O NH O Flavin adenine dinucleotide (FAD) (c) Figure 3-3 Some nucleotide derivatives. The adenosine group of each of these compounds is shown in red. Note that each also contains a vitamin derivative. (a) Coenzyme A (CoA) contains a residue of pantothenic acid (pantothenate), also known as vitamin B5. The sulfhydryl group is the site of attachment of other groups. (b) The nicotinamide group of nicotinamide adenine dinucleotide (NAD) is a derivative of the vitamin niacin (also called nicotinic acid or vitamin B3; see inset) and undergoes oxidation and reduction. The related compound nicotinamide adenine dinucleotide phosphate (NADP) contains a phosphoryl group at the adenosine C29 position. (c) Oxidation and reduction of flavin adenine dinucleotide (FAD) occurs at the riboflavin group (also known as vitamin B2). ? Identify the nitrogenous base(s) and sugar(s) in each structure. 56 Ch 3 From Genes to Proteins DNA is a double helix In a nucleic acid, the linkage between nucleotides is called a phosphodiester bond because a single phosphate group forms ester bonds to both C5′ and C3′. During DNA synthesis in a cell, when a nucleoside triphosphate is added to the polynucleotide chain, a diphosphate group is eliminated. Once incorporated into a polynucleotide, the nucleotide is formally known as a nucleotide residue. Nucleotides consecutively linked by phosphodiester bonds form a polymer in which the bases project out from a backbone of repeating sugar–phosphate groups. 5 end O O P O NH2 O N 4 H O H N H 3 N H 2 O Phosphodiester bond O N Adenine 5CH2 H P O O O N 5CH2 4 H O H N H 3 H O O H3C 5CH2 H O NH O H H H 2 O H P NH2 O O N 5CH2 4 Thymine O N 3 O NH2 N H P 4 Guanine 2 O O NH O H N Cytosine O H H H 3 2 OH H 3 end The end of the polymer that bears a phosphate group attached to C5′ is known as the 5′ end, and the end that bears a free OH group at C3′ is the 3′ end. By convention, the base sequence in a polynucleotide is read from the 5′ end (on the left) to the 3′ end (on the right). DNA contains two polynucleotide strands whose bases pair through hydrogen bonding (hydrogen bonds are discussed in Section 2-1). Two hydrogen bonds link adenine and thymine, and three hydrogen bonds link guanine and cytosine: H N H N Adenine N N O CH3 H N Thymine N N O H N Guanine O N H N N H N N H O Cytosine N N H 10.85 Å All the base pairs, which consist of a purine and a pyrimidine, have the same molecular dimensions (about 11 Å wide). Consequently, the sugar–phosphate backbones of the two strands of DNA are separated by a constant distance, regardless of whether the base pair is A:T, G:C, T:A, or C:G. Sugar–phosphate backbones A T G C T A A T C G C G T A G T T C A G C A T A Although the DNA is shown here as a ladder-like structure (left), with the two sugar–phosphate backbones as the vertical supports and the base pairs as the rungs, the two strands of DNA twist around each other to generate the familiar double helix (right). This conformation allows successive base pairs, which are essentially planar, to stack on top of each other with a center-to-center distance of only 3.4 Å. In fact, Watson and Crick derived this model for DNA not just from Chargaff 's rules but also from Rosalind Franklin's studies of the diffraction (scattering) of an X-ray beam by a DNA fiber, which suggested a helix with a repeating spacing of 3.4 Å. 57 DNA Is the Genetic Material 58 Ch 3 From Genes to Proteins 5 end 3 end Major groove 34 Å (10 bp) Minor groove 3 end 5 end 20 Å (a) (b) Figure 3-4 Model of DNA. (a) Ball-and-stick model with atoms colored: C gray, O red, N blue, and P gold (H atoms are not shown). (b) Space-filling model with the sugar–phosphate backbone in gray and the bases color-coded: A green, C blue, G yellow, and T red. See Interactive Exercise. Three-dimensional structure of DNA. ? How many nucleotides are shown in this double helix? The major features of the DNA molecule include the following (Fig. 3-4): 1. The two polynucleotide strands are antiparallel; that is, their phosphodiester bonds run in opposite directions. One strand has a 5′ S 3′ orientation, and the other has a 3′ S 5′ orientation. 2. The DNA "ladder" is twisted in a right-handed fashion. (If you climbed the DNA helix as if it were a spiral staircase, you would hold the outer railing—the sugar–phosphate backbone—with your right hand.) 3. The diameter of the helix is about 20 Å, and it completes a turn about every 10 base pairs, which corresponds to an axial distance of about 34 Å. 4. The twisting of the DNA "ladder" into a helix creates two grooves of unequal width, the major and minor grooves. 5. The sugar–phosphate backbones define the exterior of the helix and are exposed to the solvent. The negatively charged phosphate groups bind Mg2⩽ cations in vivo, which helps minimize electrostatic repulsion between these groups. 6. The base pairs are located in the center of the helix, approximately perpendicular to the helix axis. 7. The base pairs stack on top of each other, so the core of the helix is solid (see Fig. 3-4b). Although the planar faces of the base pairs are not accessible to the solvent, their edges are exposed in the major and minor grooves (this allows certain DNA-binding proteins to recognize specific bases). In nature, DNA seldom assumes a perfectly regular conformation because of small sequence-dependent irregularities. For example, base pairs can roll or twist like propeller blades, and the helix may wind more tightly or loosely at certain nucleotide sequences. DNA-binding proteins may take advantage of these small variations to locate their specific binding sites, and they in turn may further distort the DNA helix by causing it to bend or partially unwind. The size of a DNA segment is expressed in units of base pairs (bp) or kilobase pairs (1000 bp, abbreviated kb). Most naturally occurring DNA molecules comprise thousands to millions of base pairs. A short single-stranded polymer of nucleotides is usually called an oligonucleotide (oligo is Greek for "few"). In a cell, nucleotides are polymerized by the action of enzymes known as polymerases. The phosphodiester bonds linking nucleotide residues can be broken by the action of nucleases. An exonuclease removes a residue from the end of a polynucleotide chain, whereas an endonuclease cleaves at some other point along the chain. Polymerases and nucleases are usually specific for either DNA or RNA. In the absence of these enzymes, the structures of nucleic acids are remarkably stable. RNA is single-stranded RNA, which is a single-stranded polynucleotide, has greater conformational freedom than DNA, whose structure is constrained by the requirements of regular basepairing between its two strands. An RNA strand can fold back on itself so that base pairs form between complementary segments of the same strand. Consequently, RNA molecules tend to assume intricate three-dimensional shapes (Fig. 3-5). Unlike DNA, whose regular structure is suited for the long-term storage of genetic information, RNA can assume more active roles in expressing that information. For example, the molecule shown in Figure 3-4, which carries the amino acid phenylalanine, interacts with a number of proteins and other RNA molecules during protein synthesis. The residues of RNA are also capable of base-pairing with a complementary single strand of DNA to produce an RNA–DNA hybrid double helix. A double helix involving RNA is wider and flatter than the standard DNA helix (its diameter is about 26 Å, and it makes one helical turn every 11 residues). In addition, its base pairs are inclined to the helix axis by about 20° (Fig. 3-6). These structural differences relative to the standard DNA helix primarily reflect the presence of the 2′ OH groups in RNA. A double-stranded DNA helix can adopt this same helical conformation; it is known as A-DNA. The standard DNA helix shown in Figure 3-4 is known as B-DNA. Other conformations of DNA have been described, and there is evidence that they exist in vivo, at least for certain nucleotide sequences, but their functional significance is not completely understood. Figure 3-5 A transfer RNA molecule. This 76-nucleotide single-stranded RNA molecule folds back on itself so that base pairs form between complementary segments. [Structure (pdb 4TRA) determined by E. Westhoff, P. Dumas, and D. Moras.] DNA can be denatured and renatured The pairing of polynucleotide strands in a double-stranded nucleic acid is possible because bases in each strand form hydrogen bonds with complementary bases in the other strand: A is the complement of T (or U), and G is the complement of C. However, the structural stability of the DNA helix does not depend significantly on hydrogen bonding between complementary bases. (If the strands were separated, the bases could still satisfy their hydrogen-bonding requirements by forming hydrogen bonds with solvent water molecules.) Instead, stability depends mostly on stacking interactions, which are a form of van der Waals interaction, between adjacent base pairs. A view down the helix axis shows that stacked base pairs do not overlap exactly, due to the winding of the helix (Fig. 3-7). Although individual stacking interactions are weak, they are additive along the length of a DNA molecule. The stacking interactions between neighboring G:C base pairs are stronger than those of A:T base pairs (this is not related to the fact that G:C base pairs have one more hydrogen bond than A:T base pairs). Consequently, a DNA helix that is rich in G and C is harder to disrupt than DNA with a high proportion of A and T. These differences can be quantified in the melting temperature (Tm) of the DNA. Figure 3-6 An RNA–DNA hybrid helix. In a double helix formed by one strand of RNA (red) and one strand of DNA (blue), the planar base pairs are tilted and the helix does not wind as steeply as in a standard DNA double helix (compare with Fig. 3-4). [Structure (pdb 1FIX) determined by N. C. Horton and B. C. Finzel.] 59 60 Ch 3 From Genes to Proteins Relative absorbance at 260 nm pairs. A view down the central axis of the DNA helix shows the overlap of neighboring base pairs (only the first two nucleotide pairs are highlighted). ? Locate the base and sugar in the blue nucleotides. Melting ends + 1.3 1.2 1.1 1.0 Figure 3-7 Axial view of DNA base Melting begins 1.4 Tm 30 50 70 Temperature (°C) 90 Figure 3-8 A DNA melting curve. Thermal denaturation (melting, or strand separation) of DNA results in an increase in ultraviolet absorbance relative to the absorbance at 25°C. The melting point, Tm, of the DNA sample is defined as the midpoint of the melting curve. To determine the melting point of a sample of DNA, the temperature is slowly increased. At a sufficiently high temperature, the base pairs begin to unstack, hydrogen bonds break, and the two strands begin to separate. This process continues as the temperature rises, until the two strands come completely apart. The melting, or denaturation, of the DNA can be recorded as a melting curve (Fig. 3-8) by monitoring an increase in the absorbance of ultraviolet (260-nm) light (the aromatic bases absorb more light when unstacked). The midpoint of the melting curve (that is, the temperature at which half the DNA has separated into single strands) is the Tm. Table 3-2 lists the GC content and the melting point of the DNA from different species. Since manipulating DNA in the laboratory frequently requires the thermal separation of paired DNA strands, it is sometimes helpful to know the DNA's GC content. When the temperature is lowered slowly, denatured DNA can renature; that is, the separated strands can re-form a double helix by reestablishing hydrogen bonds between the complementary strands and by restacking the base pairs. The maximum rate of renaturation occurs at about 20–25°C below the melting temperature. If the DNA is cooled too rapidly, it may not fully renature because base pairs may form randomly between short complementary segments. At low temperatures, the improperly paired segments are frozen in place since they do not have enough thermal energy to melt apart and find their correct complements (Fig. 3-9). The rate of renaturation of denatured DNA depends on the length of the double-stranded molecule: Short segments come together (anneal) faster than longer segments because the bases in each strand must locate their partners along the length of the complementary strand. The ability of short single-stranded nucleic acids (either DNA or RNA) to hybridize with longer polynucleotide chains is the basis for a number of useful laboratory techniques (described in detail in Section 3-4). For example, an oligonucleotide probe TABLE 3-2 GC Content and Melting Points of DNA Source of DNA Dictyostelium discoideum (fungus) Clostridium butyricum (bacterium) Homo sapiens Streptomyces albus (bacterium) GC Content (%) Tm (8C) 23.0 37.4 40.3 72.3 79.5 82.1 86.5 100.5 [Data from Brown, T. A. (ed.), Molecular Biology LabFax, vol. I., Academic Press (1998), pp. 233–237.] Cooling to 20–25°C below Tm (renaturation) High heat (melting) Rapid cooling to temperature much lower than Tm (improper base pairing) 61 Genes Encode Proteins Rewarming to 20–25 °C below Tm (renaturation) Figure 3-9 Renaturation of DNA. DNA strands that have been melted apart can renature at a temperature of 20–25°C below the Tm. At much lower temperatures, base pairs may form between short complementary segments within and between the single strands. Correct renaturation is possible only if the sample is rewarmed so that the improperly paired strands can separate and reanneal. that has been tagged with a radioactive isotope or a fluorescent group can be used to detect the presence of a complementary nucleic acid sequence in a complex mixture. CONCEPT REVIEW • What are the relationships among purines, pyrimidines, nucleosides, nucleotides, and nucleic acids? • Describe the arrangement of the base pairs and sugar–phosphate backbones in DNA. • What did Chargaff's rules reveal about the structure of DNA? • How do DNA and RNA differ? • Describe the molecular events in DNA denaturation and renaturation. 3-2 Genes Encode Proteins The complementarity of the two strands of DNA is essential for its function as the storehouse of genetic information, since this information must be replicated (copied) for each new generation. As first suggested by Watson and Crick, the separated strands of DNA direct the synthesis of complementary strands, thereby generating two identical double-stranded molecules (Fig. 3-10). The parental strands are said to act as templates for the assembly of the new strands because their sequence of nucleotides determines the sequence of nucleotides in the new strands. Thus, genetic information—in the form of a sequence of nucleotide residues—is transmitted each time a cell divides. A similar phenomenon is responsible for the expression of that genetic information, a process in which the information is used to direct the synthesis of proteins that carry out the cell's activities. First, a portion of the DNA, a gene, is transcribed to produce a complementary strand of RNA; then the RNA is translated into protein. This paradigm, known as the central dogma of molecular biology, was formulated by Francis Crick. It can be shown schematically as lication rep transcription DNA translation RNA Protein KEY CONCEPT • The biological information encoded by a sequence of DNA is transcribed to RNA and then translated into the amino acid sequence of a protein. See Guided Exploration. Overview of transcription and translation. 62 Ch 3 From Genes to Proteins Old Old Even in the simplest organisms, DNA is an enormous molecule, and many organisms contain several different DNA molecules (for example, the chromosomes of eukaryotes). An organism's complete set of genetic information is called its genome. A genome may comprise several hundred to perhaps 35,000 genes. To transcribe a gene, one of the two strands of DNA serves as a template for an RNA polymerase to synthesize a complementary strand of RNA. The RNA therefore has the same sequence (except for the substitution of U for T) and the same 5′ S 3′ orientation as the nontemplate strand of DNA. This strand of DNA is often called the coding strand (the template strand is called the noncoding strand). A DNA 3′ New New Old Figure 3-10 DNA replication. The double helix unwinds so that each parental strand can serve as a template for the synthesis of a new complementary strand. The result is two identical double-helical DNA molecules. ? 5′ T DNA Old New A A G G T A G C C A T C Coding strand (nontemplate) G T C 5′ New G C T A RNA C U T G A G C UG A T C C G A 3′ C G C T T A C G A G C G A A T G C T 5′ Noncoding strand (template) The transcribed RNA is known as messenger RNA (mRNA) because it carries the same genetic message as the gene. The mRNA is translated in the ribosome, a cellular particle consisting of protein and ribosomal RNA (rRNA). At the ribosome, small molecules called transfer RNA (tRNA), which carry amino acids, recognize sequential sets of three bases (known as codons) in the mRNA through complementary base-pairing (a tRNA molecule is shown in Fig. 3-5). The ribosome covalently links the amino acids carried by successive tRNAs to form a protein. The protein's amino acid sequence therefore ultimately depends on the nucleotide sequence of the DNA. 5′ DNA 3′ Label the 5′ and 3′ end of each strand. C T C A G T G C C G A G T C A C G G 3′ 5′ Transcription mRNA 5′ tRNAs C U C A G U G C C G A G U C A C G G Leu Ser 3′ Ala Translation Protein Leucine Serine Alanine The correspondence between amino acids and mRNA codons is known as the genetic code. There are a total of 64 codons: 3 of these are "stop" signals that ter- minate translation, and the remaining 61 represent, with some redundancy, the 20 standard amino acids found in proteins. Table 3-3 shows which codons specify which amino acids. In theory, knowing a gene's nucleotide sequence should be equivalent to knowing the amino acid sequence of the protein encoded by the gene. However, as we will see, genetic information is often "processed" at several points before the protein reaches its mature form. Keep in mind that the rRNA and tRNA required for protein synthesis, as well as other types of RNA, are also encoded by genes. The "products" of these genes are the result of transcription without translation. TABLE 3-3 First Position (59 end) 63 Genes Encode Proteins The Standard Genetic Codea Third Position (39end) Second Position U C A G U UUU UUC UUA UUG Phe Phe Leu Leu UCU UCC UCA UCG Ser Ser Ser Ser UAU UAC UAA UAG Tyr Tyr Stop Stop UGU UGC UGA UGG Cys Cys Stop Trp U C A G C CUU CUC CUA CUG Leu Leu Leu Leu CCU CCC CCA CCG Pro Pro Pro Pro CAU CAC CAA CAG His His Gln Gln CGU CGC CGA CGG Arg Arg Arg Arg U C A G A AUU AUC AUA AUG Ile Ile Ile Met ACU ACC ACA ACG Thr Thr Thr Thr AAU AAC AAA AAG Asn Asn Lys Lys AGU AGC AGA AGG Ser Ser Arg Arg U C A G G GUU GUC GUA GUG Val Val Val Val GCU GCC GCA GCG Ala Ala Ala Ala GAU GAC GAA GAG Asp Asp Glu Glu GGU GGC GGA GGG Gly Gly Gly Gly U C A G a The 20 amino acids are abbreviated; Ala, alanine; Arg, arginine; Asn, asparagine; Asp, aspartate; Cys, cysteine; Gly, glycine; Gln, glutamine; Glu, glutamate; His, histidine; Ile, isoleucine; Leu, leucine; Lys, lysine; Met, methionine; Phe, phenylalanine; Pro, proline; Ser, serine; Thr, threonine; Trp, tryptophan; Tyr, tyrosine; and Val, valine. ? How many amino acids would be uniquely specified by a genetic code that consisted of just the first two nucleotides in each codon? A mutated gene can cause disease Because an organism's genetic material influences the organism's entire repertoire of activities, it is vitally important to unravel the sequence of nucleotides in that organism's DNA, even by examining one gene at a time. Thousands of genes have been identified through studies of the genes' protein products, and millions more have been catalogued through genome-sequencing projects (discussed below in Section 3-3). Although the functions of many genes are not yet understood, some genes have come to light through the study of inherited diseases. In a traditional approach, researchers have used the defective protein associated with a particular disease to track down the relevant genetic defect. For example, the variant hemoglobin protein that causes sickle cell anemia results from the substitution of the amino acid glutamate by valine. In the gene for that protein chain, the normal GAG codon has been mutated (altered) to GTG. Normal gene Protein · · · ACT CCT GAG GAG AAG · · · · · · Thr – Pro – Glu – Glu – Lys · · · Mutated gene · · · ACT CCT GTG GAG AAG · · · Protein · · · Thr – Pro – Val – Glu – Lys · · · More modern approaches begin with analysis of the DNA to discover the genetic changes that lead to disease. The first successful application of this method identified the cystic fibrosis gene, that is, the gene whose mutation cause the disease (Box 3-A). Over 3000 genes have been linked to specific monogenetic diseases, such as sickle cell anemia and cystic fibrosis. In many cases, a variety of different mutations have been catalogued for each disease gene, which explains in part why symptoms of the disease vary between individuals. The database known as OMIM (Online Mendelian Inheritance in Man; contains information on thousands of genetic variants, including the clinical features of the resulting disorder and its biochemical basis. The Genetic Testing Registry (is a database of the diseases that can be detected through analysis of DNA, carried out by either clinical or research laboratories. BOX 3-A CLINICAL CONNECTION Discovery of the Cystic Fibrosis Gene About 1 in 3000 babies in the United States is born with cystic fibrosis (CF), DNA the most common inherited disease in individuals of northern European extraction. The most serious symptom of CF is the obstruction of the airways by thick, Unspliced sticky mucus, which tends to create an mRNA ideal environment for bacterial growth. Individuals with CF may also suffer from impaired secretion of digestive enzymes from the pancreas, which contributes to malnutrition and poor growth. Histori- Spliced mRNA cally, individuals with CF died in childhood, but a variety of treatments, including the use of antibiotics to prevent lung infections, have now extended survival well into adulthood. Before the era of DNA testing, one of the diagnostic signs of CF was high chloProtein ride concentrations in sweat (according to medieval folklore, a baby who tasted salty when kissed was predicted to die soon). But neither this characteristic nor other symptoms, such as the thick mucus in the airways, pointed unequivocally to a defect in any known protein. Consequently, the search for the genetic basis for CF required a strategy that did not use a protein as its starting point. To find the cystic fibrosis gene, researchers analyzed DNA from affected individuals, who had two copies of the defective CF gene, and from family members who were asymptomatic Normal gene carriers and had one normal and one defective Protein copy of the gene. Individuals with one or two copies of the defective CF gene shared two other genetic features that can be detected in a laboratory test. These two DNA markers were used to Mutated gene define a chromosomal region that was likely to Protein contain the cystic fibrosis gene. In particular, one DNA segment on chromosome 7 appeared to be present in a number of mammalian species, which suggested that the segment contained an essential gene (about 98% of mammalian DNA does not encode any protein). The researchers then deduced the sequence of nucleotides in this region of DNA, ultimately identifying a stretch of about 250,000 bp as the CF gene. As is the case for nearly all mammalian genes, only certain portions of the CF gene directly correspond to a protein product, because segments of the mRNA molecule transcribed from the gene are excised, an event called splicing, before the message is translated into protein (splicing is discussed further in Section 21-3). In addition, sequences at each end of the mRNA are not translated. After splicing, the mRNA is only 6129 nucleotides long. Of this molecule, 4440 nucleotides (or 4440 4 3 5 1480 codons) specify the 1480 amino acid residues of the protein product. 250,000 bp transcription splicing 6129 bases 4440 bases 4440 bases = 1480 codons translation 1480 amino acids Matching every three bases in the derived mRNA sequence with the appropriate amino acid (see Table 3-3) yielded the amino acid sequence of the protein. Additional sequencing studies showed that in about 70% of CF patients, the gene is missing three nucleotides. This results in the deletion of a single phenylalanine (Phe) residue at position 508 (the 508th amino acid residue in the encoded protein): 504 505 506 507 508 509 510 511 512 . . . GAA AAT ATC ATC TTT GGT GTT TCC TAT . . . . . . Glu Asn Ile Ile Phe Gly Val Ser Tyr . . . 504 505 506 507 508 509 510 511 512 . . . GAA AAT ATC AT– – –T GGT GTT TCC TAT . . . . . . Glu Asn Ile Ile Gly Val Ser Tyr . . . Note that although the nucleotide deletion affects codons 507 and 508, the redundancy of the genetic code means that the isoleucine (Ile) at position 507 is not affected (because codons ATC and ATT both specify Ile). The protein lacking Phe 508 is abnormally processed by the cell, so very little is present in functional form. So what does the CF gene do? The putative function of the cystic fibrosis gene product was identified by its sequence similarity to a large family of proteins involved in the transport of substances across cell membranes (recall from Section 2-2 that only nonpolar substances can spontaneously traverse a lipid bilayer; all other substances require a protein transporter). Each member of this protein family has one or two segments that position the protein in the membrane. The CF protein also contains an additional domain thought to play a regulatory role. Accordingly, the protein was named the cystic fibrosis transmembrane conductance regulator (CFTR). When the CFTR gene was introduced into different cell types, its function could be studied. The CFTR protein is, in fact, a membrane protein that acts as a channel to allow Cl2 to exit the cell EXTRACELLULAR FLUID Cl − CYTOSOL Cl − Cell membrane CFTR CFTR also appears to regulate Na1 uptake by the cell. Consequently, a defective or absent CFTR protein disrupts the normal distribution of Na1 and Cl2. In the CF lung, the concentrations of the ions are low in the extracellular space. As a result, the water that would normally be drawn by high concentrations of these ions is absent. In a normal lung, the extracellular fluid is thin and watery, but in the CF lung, the fluid is thick and viscous. In the sweat gland, a defective CFTR alters the transport of Na1 and Cl2, causing the salty sweat that is diagnostic of CF. Questions: 1. The Phe-deletion mutation described on the facing page causes a severe form of CF. Other types of mutations in the CF gene produce milder forms of the disease that may not be detected until adulthood. Explain. 2. One portion of the normal CF gene has the sequence . . . AAT ATA GAT ACA G . . . In some individuals with cystic fibrosis, this portion of the gene has the sequence . . . AAT AGA TAC AG . . . How has the DNA sequence changed and how does this affect the encoded protein? 3. Most genetic diseases that limit survival to reproductive age are relatively rare; CF is an exception. The prevalence of the defective CF gene suggests that it may have had benefits during human evolution. Some pathogens (disease-causing organisms) use the normal CFTR as an entry point for infecting cells. Explain why individuals with one normal and one defective CF gene are more likely to survive (and pass on their genes) than individuals with two normal or two defective copies of the CF gene. 4. Would cystic fibrosis be a good candidate for treatment by gene therapy (Section 3-4)? CONCEPT REVIEW • How does DNA encode genetic information and how is this information expressed? • What is the relationship between the nucleotide sequence in a gene and the amino acid sequence of a protein? • List some reasons why knowing a gene's sequence might be useful. 3-3 Genomics The ability to sequence large tracts of DNA has made it possible to study entire genomes, from the small DNA molecules of parasitic bacteria to the enormous multichromosome genomes of plants and mammals. Sequence data are customarily deposited in a public database such as GenBank. The data can be accessed electronically in order to compare a given sequence to sequences from other genes (see Bioinformatics Project 2). Some of the thousands of organisms whose genomes have been partially or fully sequenced are listed in Table 3-4. This list includes species that are widely used as model organisms for different types of biochemical studies (Fig. 3-11). KEY CONCEPTS • The genomes of different species vary in size and number of genes. • Genes can be identified by their nucleotide sequences. • Analysis of genetic data can provide information about gene function and risk of disease. Gene number is roughly correlated with organismal complexity Not surprisingly, organisms with the simplest lifestyles tend to have the least amount of DNA and the fewest genes. For example, M. genitalium and H. influenzae (see Table 3-4) are human parasites that depend on their host to provide nutrients; these organisms do not contain as many genes as free-living bacteria such as Synechocystis (a photosynthetic bacterium). Multicellular organisms generally have even more DNA and more genes, presumably to support the activities of their many specialized 65 66 Ch 3 From Genes to Proteins TABLE 3-4 Genome Size and Gene Number of Some Organisms Organism Genome Size (kb) Number of Genes Bacteria Mycoplasma genitalium Haemophilus influenzae Synechocystis PCC6803 Escherichia coli 580 1,830 3,947 4,643 525 1,789 3,618 4,630 Archaea Methanocaldococcus jannaschii Archaeoglobus fulgidus 1,740 2,178 1,830 2,486 Fungi Saccharomyces cerevisiae (yeast) 12,071 6,281 119,146 382,151 ≈2,046,000 33,323 30,294 ≈32,000 100,268 139,466 3,102,000 21,175 15,016 ≈21,000 Plants Arabidopsis thaliana Oryza sativa (rice) Zea mays (corn) Animals Caenorhabditis elegans (nematode) Drosophila melanogaster (fruit fly) Homo sapiens [Data from NCBI Genome Project.] ? What is the relationship between genome size and gene number in prokaryotes? How does this differ in eukaryotes? Figure 3-11 Some model organisms. (a) Escherichia coli, a normal inhabitant of the mammalian digestive tract, is a metabolically versatile bacterium that tolerates both aerobic and anaerobic conditions. (b) Baker's yeast, Saccharomyces cerevisiae, is one of the simplest eukaryotic organisms, with just over 6000 genes. (c) Caenorhabditis elegans is a small (1-mm) and transparent roundworm. As a multicellular organism, it bears genes not found in unicellular organisms. (d) The plant kingdom is represented by Arabidopsis thaliana, which has a short generation time and readily takes up foreign DNA. [Dr. Kari Lounatmaa/Science (a) (b) (c) (d) Photo Library/Photo Researchers; Andrew Syred/Science Photo Library/Photo Researchers; Sinclair Stammers/Science Photo Library/Photo Researchers; Dr. Jeremy Burgess/Science Photo Library/Photo Researchers.] Unique sequence Moderately repetitive Protein-coding Highly repetitive Figure 3-12 Coding and noncoding portions of the human genome. Approximately 1.5% of the genome codes for proteins. Moderately repetitive sequences make up 45% of the genome and highly repetitive sequences about 3%, so that roughly half of the human genome consists of unique DNA sequences of unknown function. Up to 80% of the genome may be transcribed, however. cell types. Interestingly, humans contain about as many genes as nematodes, suggesting that organismal complexity results not just from the raw number of genes but from how the genes are transcribed and translated into protein. Note that humans and many other organisms are diploid (having two sets of genetic information, one from each parent), so that each human cell contains roughly 6.2 billion base pairs of DNA. For simplicity, genetic information usually refers to the haploid state, equivalent to one set of genetic instructions. In prokaryotic genomes, all but a few percent of the DNA represents genes for proteins and RNA. The proportion of noncoding DNA generally increases with the complexity of the organism. For example, about 30% of the yeast genome, about half of the Arabidopsis genome, and over 98% of the human genome is noncoding DNA. Although up to 80% of the human genome may actually be transcribed to RNA, the protein-coding segments account for only about 1.5% of the total (Fig. 3-12). Much of the noncoding DNA consists of repeating sequences with no known function. The presence of repetitive DNA helps explain why certain very large genomes actually include only a modest number of genes. For example, the maize (corn) and rice genomes contain about the same number of genes, but the maize genome is as much as 10 times larger than the rice genome. Over half of the maize genome appears to be composed of transposable elements, short segments of DNA that are copied many times and inserted randomly into the chromosomes. The human genome contains several types of repetitive DNA sequences, including the inactive remnants of transposable elements. About 45% of human DNA consists of moderately repetitive sequences, which are blocks of hundreds or thousands of nucleotides scattered throughout the genome. The most numerous of these are present in hundreds of thousands of copies. Highly repetitive sequences account for another 3% of the human genome. These segments of 2 to 10 bases are present in millions of copies. They are repeated tandemly (side by side), sometimes thousands of times. The number of repeats of a given sequence often varies between individuals, even in the same family, so this information can be analyzed to produce a DNA "fingerprint" (see Section 3-4). Some so-called noncoding DNA in fact consists of genes for RNA molecules, which appear to play a variety of roles in regulating the expression of protein-coding genes (Chapter 21). Comparisons of mammalian genomes indicate that as much as 6% of the human genome appears to have changed relatively little during evolution, suggesting that it has some essential function. Genes are identified by comparing sequences For many genomes, the exact number of genes has not yet been determined, and different methods for identifying genes yield different estimates. For example, a 67 Genomics 68 Ch 3 From Genes to Proteins (a) (b) Figure 3-13 Examples of genome maps. (a) Genes located in a 10-kb span of the E. coli chromosome. (b) Genes from a 2500-kb gene-rich segment of the mouse genome. Each gene is represented by a colored block. computer can scan a DNA sequence for an open reading frame (ORF), that is, a stretch of nucleotides that can potentially be transcribed or translated. For a proteincoding gene, the ORF begins with a "start" codon: ATG in the coding strand of DNA, which corresponds to AUG in RNA (see Table 3-3). This codon specifies methionine, the initial residue of all newly synthesized proteins. The ORF ends with one of the three "stop" codons: DNA coding sequences of TAA, TAG, or TGA, which correspond to the three mRNA stop codons (see Table 3-3). Other so-called ab initio ("from the beginning") gene-identifying methods scan the DNA for other features that characterize the beginnings and endings of genes. These methods tend to overestimate the number of genes. Another method for identifying genes in a genome relies on sequence comparisons with known genes (and thereby probably underestimates the true number of genes). Such genome-to-genome comparisons are possible because of the universal nature of the genetic code and the relatedness of all organisms through evolution (Section 1-4). Genes with similar functions in different species tend to have similar sequences; such genes are said to be homologous. Even an inexact match can still indicate a protein's functional category, such as enzyme or hormone receptor, although its exact role in the cell may not be obvious. Genes that appear to lack counterparts in other species are known as orphan genes. At present, the number of known genes exceeds the number of known gene products (proteins and RNA molecules). This is hardly surprising, given that some genes are expressed at low levels or generate products that have not yet been detected through conventional biochemical isolation approaches. About 20% of the genes in the well-studied organism E. coli have not yet been assigned functions. Genome maps, such as the ones shown in Figure 3-13, indicate the placement and orientation of genes on a chromosome. Arrows pointing in opposite directions represent genes encoded by different strands of the double-stranded chromosome. Note that mammalian genes are typically much longer than bacterial genes (27 kb on average), since they contain sequences that are spliced out of the transcript before translation. In addition, the spaces between genes are much larger in the mammalian genome. Gene-mapping projects have uncovered some interesting aspects of evolution, including horizontal gene transfer. This occurs when a gene is transferred between species rather than from parent to offspring of the same species (vertical gene transfer). Horizontal gene transfer may be mediated by viruses, which can pick up extra DNA as they insert and excise themselves from the host's chromosomes. This activity can generate, for example, what appears to be a mammalian gene inside a bacterial genome. The ease with which many bacterial organisms trade their genes has given rise to the idea that groups of bacteria should be viewed as a continuum of genomic variations instead of separate species with discrete genomes. Genomic data can be linked to disease Genomics, the study of genomes, has a number of practical applications. For one thing, the number of genes and their putative functions provide a rough snapshot of the (a) Human (b) Drosophila cell communication immune system Figure 3-14 Functional classification of genes. This diagram is based on 17,181 human genes (a) and 9837 Drosophila genes (b), grouped according to the biochemical function of the gene product. Humans devote a larger proportion of genes to cell communication (25.4% versus 18.6% in Drosophila) and to the immune system (15.3% versus 10.2% in Drosophila). [Data from the Protein Analysis through Evolutionary Relationships classification system, www.pantherdb.org/.] metabolic capabilities of a given organism. For example, humans and fruit flies differ in the number of genes that code for cell-signaling pathways and immune system functions (Fig. 3-14). An unusual number of genes belonging to one category might indicate some unusual biological property in an organism. This sort of knowledge could be useful for developing drugs to inhibit the growth of a pathogenic organism according to its unique metabolism. Genomic analysis also reveals variations among individuals, some of which can be linked to an individual's chance of developing a particular disease. In addition to genetic changes that are clearly associated with a single-gene disorder, millions more sequence variations have been catalogued. On average, the DNA of any two humans differs at 3 million sites, or about once every thousand base pairs. These singlenucleotide polymorphisms (SNPs, instances where the DNA sequence differs among individuals) are compiled in databases. Some of the factors that can alter DNA are discussed in Section 20-5). Researchers have attempted to correlate SNPs with disorders, such as cardiovascular disease or cancer, that likely depend on the contributions of many genes. Genome-wide association studies (GWAS) have identified, for example, 39 sites that are associated with type 2 diabetes and 71 that are associated with Crohn's disease, an autoimmune disorder. The risk tied to any particular genetic variant is low, but the entire set of variations can explain up to 50% of the heritability of the disease. Although the SNPs are only proxies for disease genes, these data should provide a starting point for researchers to explore the DNA near the SNPs to discover the genes that are directly involved in the disease. Several commercial enterprises offer individual genome-sequencing services, but until genetic information can be reliably translated into effective disease-prevention or treatment regimens, the practical value of "personal genomics" is somewhat limited. CONCEPT REVIEW • Describe the rough correlation between gene number and organismal lifestyle. • How are genes identified? • Why is it useful to identify homologs of human genes? • What is the value of genome-wide association studies? What are their limitations? 69 Genomics 3-4 Tools and Techniques: Manipulating DNA KEY CONCEPTS • A DNA molecule can be sequenced or amplified by using DNA polymerase to make a copy of a template strand. • Linking together DNA fragments produces recombinant DNA molecules that can be used to study gene function, to express genes in host organisms, and to engineer genes for commercial and therapeutic purposes. See Guided Exploration. DNA sequence determination by the chainterminator method. Molecular biologists have devised clever procedures for isolating, amplifying, sequencing, altering, and otherwise manipulating DNA. Many of the techniques take advantage of naturally occurring enzymes that cut, copy, and link nucleic acids. These techniques also exploit the ability of nucleic acids to interact with complementary molecules purified from natural sources or synthesized in the laboratory. DNA sequencing uses DNA polymerase to make a complementary strand The most widely used technique for determining the sequence of nucleotides in a segment of DNA was developed by Frederick Sanger in 1975. The Sanger sequencing technique is also known as dideoxy DNA sequencing because it uses dideoxy nucleotides, that is, nucleotides lacking both 29 and 39 hydroxyl groups: P P P base OCH2 O H H H H H H 2,3-Dideoxynucleoside triphosphate (ddNTP) HOW do researchers decipher the information in DNA? 70 A dideoxynucleoside triphosphate can be abbreviated ddNTP, where N represents any of the four bases. The sequencing protocol is outlined in Figure 3-15. First, the DNA is denatured to separate its two strands. The DNA is then incubated with a mixture of all four deoxynucleoside triphosphates (dATP, dCTP, dGTP, and dTTP) and a bacterial DNA polymerase, an enzyme that catalyzes the polymerization of the nucleotides in the order determined by their base-pairing with a single-stranded DNA template. Because the DNA polymerase cannot begin a new nucleotide strand but can only extend a preexisting chain, a short single-stranded primer that base pairs with the template strand is added to the mixture. Keep in mind that the reaction mixture actually contains millions of molecules of the template, the primer, and the polymerase. The reaction mixture also includes small amounts of four dideoxynucleotides (ddATP, ddCTP, ddGTP, and ddTTP), each of which is tagged with a different fluorescent dye. As the DNA polymerase proceeds to synthesize a new DNA chain, it occasionally adds one of the ddNTPs in place of the corresponding dNTP. This halts further extension of the DNA chain because the newly incorporated ddNTP, which lacks a 3′ OH group, cannot form the 3′ portion of a 3′–5′ phosphodiester bond to the next nucleotide (see Section 3-1). The concentrations of the ddNTPs in the reaction mixture are lower than the concentrations of their corresponding dNTPs, so the DNA polymerase can assemble new chains of varying length before the random incorporation of a ddNTP halts further polymerization. The result is a population of truncated DNA strands, each capped by a fluorescent ddNTP residue. The fragments differ in length by a single nucleotide since they all started from identical primers. The reaction products are subjected to electrophoresis, a procedure in which the molecules move through a gel-like matrix such as agarose or polyacrylamide under the influence of an electric field. Because all the DNA segments have a uniform charge density, they are separated on the basis of their size (the smallest molecules move fastest). The separated molecules are excited by a laser so that the fluorescent dye 5′ — G T C A T A GC T C G A CG T A G— 3′ 3′ — C A G T A T CG A GC T GC A T C— 5′ Double-stranded DNA 71 Tools and Techniques: Manipulating DNA Denature DNA 3′ — C A G T A T CG A GC T GC A T C— 5′ Single DNA strand to be sequenced Add primer, DNA polymerase, dNTPs, and fluorescent-labeled ddNTPs Primer 5′ A T A GC 3′ — C A G T A T CG A GC T GC A T C— 5′ Primer bound to DNA DNA polymerase extends the primer by sequentially adding nucleotides that base pair with the template strand. 5′ - A T A GC T CG A C 3′ — C A G T A T CG A GC T GC A T C— 5′ Template strand The occasional incorporation of a ddNTP terminates polymerization. ATA G C TC G A C G TA G ATA G C TC G A C G TA Set of newly synthesized DNA chains differing by one nucleotide and terminating with a labeled ATA G C TC G A C G T dideoxynucleotide ATA G C TC G A C G ATA G C TC G A C ATA G C TC G A Fluorescent residue ATA G C TC G ATA G C TC ATA G C T Primer Figure 3-15 Dideoxy DNA sequencing procedure. attached to each dideoxynucleotide residue emits its characteristic color (Fig. 3-16). The order of appearance of the colors corresponds to the order of nucleoides in the newly synthesized DNA. Keep in mind that this sequence is complementary to the DNA strand that was used as a template in the sequencing reaction. With automation, a single reaction can yield a sequence of 400–1000 nucleotides. Newer sequencing procedures also use DNA polymerase to generate a complementary DNA strand. In the pyrosequencing technique, molecules of the template DNA are immobilized on a plastic surface, a primer and DNA polymerase are added, A AT G C C TAT T G G AT C C A Fluorescence detector Capillary tube Migration of DNA fragments – Laser light + Figure 3-16 Results of dideoxy sequencing. In electrophoresis, the DNA fragments generated by dideoxy sequencing move through a thin tube under an electric field and are detected by their fluorescence. The peaks represent fragments that differ in size by one nucleotide, and the color corresponds to the dideoxynucleotide at the end of the fragment. Thus, the successive peaks reveal the nucleotide sequence. 72 Ch 3 From Genes to Proteins and then a dNTP substrate is introduced. If DNA polymerase adds that nucleotide to the new DNA strand, pyrophosphate (the "diphosphate" portion of the dNTP) is released and triggers a chemical reaction involving the firefly enzyme luciferase, which generates a flash of light. Solutions of each of the four dNTPs are successively washed across the immobilized DNA template, and a detector records whether light is produced in the presence of a particular dNTP. In this way, the sequence of nucleotides complementary to the template strand can be deduced. In a similar approach, the DNA to be sequenced is introduced into microscopic wells positioned atop an ionsensitive layer. When DNA polymerase adds the appropriate nucleotide, the reaction releases a hydrogen ion, which is detected as a voltage change. Pyrosequencing and related techniques can "read" stretches of 300–500 nucleotides, somewhat shorter than the sequences revealed by dideoxy sequencing, but the small scale of the fluiddelivery system allows many templates to be sequenced simultaneously. Because DNA samples are frequently too large to be sequenced in a single pass, the DNA must be broken into a number of overlapping segments that are individually sequenced. The nucleotide sequence of the entire DNA can then be reconstructed by computer analysis of the overlapping sequences. Although this piece-by-piece approach can be carried out systematically, it is usually more efficient to generate a large number of DNA fragments in a random, or "shotgun," fashion and then examine the sequenced pieces for overlaps. Original DNA Segments to be individually sequenced The polymerase chain reaction amplifies DNA See Guided Exploration. PCR and site-directed mutagenesis. At one time, scientists requiring large amounts of a particular DNA sequence had to laboriously isolate the target DNA and clone (copy) it in living cells, as described later in this section. That changed in 1985 with the invention of the polymerase chain reaction (PCR) by Kary Mullis. Although PCR cannot entirely replace cloning as a research tool, it provides a relatively easy and rapid way to amplify a segment of DNA. One of the advantages of PCR over traditional cloning techniques is that the starting material need not be pure (this makes the technique ideal for analyzing complex mixtures such as tissues or biological fluids). Like DNA sequencing, PCR uses DNA polymerase to make copies of a particular DNA sequence (Fig. 3-17). The reaction mixture contains the DNA sample, a DNA polymerase, all four deoxynucleotides, and two oligonucleotide primers that are complementary to the 3′ ends of the two strands of the target sequence. In the first step of PCR, the sample is heated to 90–95°C to separate the DNA strands. Next, the temperature is lowered to about 55°C, cool enough for the primers to hybridize with the DNA strands. The temperature is then increased to about 75°C, and the DNA polymerase synthesizes new DNA strands by extending the primers. The three steps—strand separation, primer binding, and primer extension—are repeated as many as 40 times. Because the primers represent the two ends of the target DNA, this sequence is preferentially amplified so that it doubles in concentration with each reaction cycle. For example, 20 cycles of PCR can theoretically yield 220 1,048,576 copies of the target sequence in a matter of hours.

The DNA can then be cloned, sequenced, or used for another purpose. One of the keys to the success of PCR is the use of bacterial DNA polymerases that can withstand the high temperatures required for strand separation (these temperatures inactivate most enzymes). Commercial PCR kits usually contain DNA polymerase from Thermus aquaticus (“Taq,” which lives in hot springs) or Pyrococcus furiosus (“Pfu,” which inhabits geothermally heated marine sediments), since their enzymes perform optimally at high temperatures. 73 Tools and Techniques: Manipulating DNA 3′ 5′ Target sequence separate strands Primers anneal primers Cycle 1 extend primers separate strands anneal primers Cycle 2 extend primers separate strands anneal primers Cycle 3 extend primers Figure 3-17 The polymerase chain reaction. Each cycle consists of separation of DNA strands, binding of primers to the 3′ ends of the target sequence, and extension of the primers by DNA polymerase. The target DNA doubles in concentration with each cycle. ? One of the limitations of PCR is that choosing appropriate primers requires some knowledge of the DNA sequence to be amplified; otherwise, the primers will not anneal with complementary sequences in the DNA and the polymerase will not be able to make any new DNA strands. However, since no new DNA is synthesized unless the primers can bind to the DNA sequence, PCR can be used to verify the presence of that sequence. In fact, PCR is the most efficient way to detect certain organisms that are difficult to grow in the laboratory. For obvious reasons, PCR is also preferred over culturing as a way to detect the presence of highly contagious bacteria and viruses. Blood banks use PCR to test donations for the human immunodeficiency virus (HIV), hepatitis viruses, and West Nile virus. Scientists have used PCR to amplify DNA from ancient bones and, in the forensics laboratory, to analyze DNA of more recent origin (Box 3-B). If merely detecting a specific DNA sequence does not provide enough information, quantitative PCR (qPCR, also known as real-time PCR) can be employed. Indicate the temperature at which each step takes place. BOX 3-B BIOCHEMISTRY NOTE DNA Fingerprinting Individuals can be distinguished by examining polymorphisms in their DNA. Current DNA fingerprinting methods use PCR to examine segments of repetitive DNA sequences, most often short tandem repeats of four nucleotides. The exact number of repeats varies among individuals, and each set of repeats, or allele, is small enough (usually less than 500 bp) that alleles differing by just one four-residue repeat can be easily differentiated. Because the first step of fingerprinting is PCR, only a tiny amount of DNA is needed—about 1 mg, or the amount present on a coffee cup or a licked envelope. And since the target segment is short, the purity and integrity of the DNA sample is usually not an issue. The locus, or region of DNA containing the repeats, is PCR-amplified using fluorescent primers that are complementary to the unique (nonrepeating) sequences flanking the repeats. In the example below, the two DNA segments have seven and eight tandem repeats of the AATG sequence. AATG AATG AATG AATG AATG AATG AATG AATG AATG AATG AATG AATG AATG AATG AATG The PCR primers hybridize with sequences flanking the repeats (shaded blue), which are the same in all individuals. The amplified products are then separated according to size by electrophoresis and detected by their fluorescence. The results are compared to reference standards containing a known number of AATG repeats, from 5 to 10 in this example. Sample A comes from an individual with two copies of the 7-repeat allele and sample B from an individual with one copy of the 7-repeat allele and one copy of the 8-repeat allele (recall that humans are diploid, with two copies of each “gene”). Each of the loci that have been selected for forensic use generally have 7 to 30 different alleles. In a single sample of DNA, multiple loci can be amplified by PCR simultaneously, proSample A vided that the sizes of the PCR products are sufficiently different that they will not overlap during electrophoresis. Alternatively, each PCR primer can bear a different fluorescent dye. Sample B The probability of two individuals having matching DNA fingerprints depends on the number of loci examined and the number of possible alleles at each locus. For example, assume that Reference one locus has 20 alleles and that each allele has a frequency in the population of 5% (1 in 20, or 1/20). Another locus has 10 alleles, and each has a frequency of 10% (1 in 10, or 1/10). The probability that two individuals would match 5 6 7 8 9 10 at both sites is 1/20 3 1/10 5 1/200 Size (the probabilities of independent events are multiplied). If multiple loci are examined, the probabilities can reach the range of 1 in a million or more. For this reason, most courts now consider DNA sequences to be unambiguous identifiers of individuals. Fluorescence 74 Ch 3 From Genes to Proteins Question: Would a child’s DNA fingerprint match the parents’ DNA fingerprints? In this technique, the polymerase chain reaction continually generates new DNA sequences that bind to fluorescent probes, so the amount of the DNA sequence can be monitored over time (rather than at the end of many reaction cycles, as in standard PCR). This approach is useful for quantifying infectious agents such as bacteria and viruses. Analysis of fetal DNA from the mother’s blood relies on the accuracy of qPCR to detect the small increase in concentration of a particular stretch of DNA that the fetus may have inherited from its mother, since circulating maternal DNA is about 10 times more abundant than the circulating fetal DNA. qPCR methods are also used to assess the level of gene expression in cells: Cellular mRNA is first reverse-transcribed to DNA, then a specific DNA sequence is amplified by PCR. A gene’s level of expression is sometimes reported relative to that of a gene that encodes a protein such as actin, which is typically produced at constant levels in cells and can therefore serve as a sort of benchmark. Restriction enzymes cut DNA at specific sequences Naturally occurring DNA molecules tend to break from mechanical stress during laboratory manipulations. However, such randomly fragmented DNA is not always useful, so researchers take advantage of enzymes that cut DNA in defined ways. Bacteria produce DNA-cleaving enzymes known as restriction endonucleases (or restriction enzymes) that catalyze the breakage of phosphodiester bonds at or near specific nucleotide sequences. These enzymes can thereby destroy foreign DNA that enters the cell (for example, bacteriophage DNA). In this way, the bacterial cell “restricts” the growth of the phage. The bacterial cell protects its own DNA from endonucleolytic digestion by methylating it (adding a OCH3 group) at the same sites recognized by its restriction endonucleases. In the laboratory, the most useful restriction enzymes are those that cleave at (rather than just near) the recognition site. Hundreds of these enzymes have been examined; some are listed in Table 3-5 along with their recognition sequences and cleavage sites. Restriction enzymes typically recognize a 4- to 8-base sequence that is identical, when read in the same 5′ S 3′ direction, on both strands. DNA with this form of symmetry is said to be palindromic (words such as madam and noon are palindromes). One restriction enzyme isolated from E. coli is known as EcoRI (the first three letters are derived from the genus and species names). Its recognition sequence is 75 Tools and Techniques: Manipulating DNA TABLE 3-5 Recognition and Cleavage Sites of Some Restriction Endonucleases Enzyme Recognition/ Cleavage Sitea AluI MspI AsuI EcoRI EcoRV PstI SauI NotI AG | CT C | CGG G | GNCCb G | AATTC GAT | ATC CTGCA | G CC | TNAGG GC | GGCCGC a The sequence of one of the two DNA strands is shown. The vertical bar indicates the cleavage site. b N represents any nucleotide. [An exhaustive source of information on restriction enzymes is available through the Restriction Enzyme Database: rebase.neb.com/rebase/rebase.html.] 23130 9416 6557 4361 5′ — G A A T T C — 3′ 3′ — C T T A A G — 5′ The arrows indicate the phosphodiester bonds that are cleaved. Note that the sequence reads the same on both strands. Because the EcoRI cleavage sites are symmetrical but staggered, the enzyme generates DNA fragments with single-stranded extensions known as sticky ends: In contrast, the E. coli restriction enzyme known as EcoRV cleaves both strands of DNA at the center of its 6-bp recognition sequence, so that the resulting DNA fragments have blunt ends: — GA T — CTA Figure 3-18 Digestion of bacteriophage DNA by the restriction enzyme HindIII. The restriction enzyme —G A A TTC — — CTT A A G— 5′ — G A T A T C — 3′ 3′ — C T A T A G — 5′ 2322 2027 ATC — TAG— Restriction enzymes have many uses in the laboratory. For example, they are indispensable for reproducibly breaking large pieces of DNA into smaller pieces of manageable size. Restriction digests of well-characterized DNA molecules, such as the 48,502 bp E. coli bacteriophage l, yield restriction fragments of predictable size (Fig. 3-18). cleaves the DNA to produce eight fragments of defined size, six of which are large enough to be separated by electrophoresis in an agarose gel. The DNA was applied to the top of the gel, and the fragments were visualized by binding a fluorescent dye. The numbers indicate the number of base pairs in each fragment. [Reprinted from www. neb.com, ; © 2012 with permission from New England Biolabs.] ? Explain why the bands at the top appear brighter than the bands at the bottom. 76 Ch 3 From Genes to Proteins DNA fragments are joined to produce recombinant DNA Recombinant DNA technology (also known as genetic engineering or molecular cloning) is a set of methods for cleaving, joining, and copying DNA fragments in vivo (PCR copies DNA in vitro). Briefly, a fragment of DNA is combined with another DNA molecule to produce a recombinant DNA molecule: 1. A fragment of DNA of the appropriate size is generated by the action of a restriction enzyme, by PCR, or by chemical synthesis. 2. The fragment is incorporated into another DNA molecule. 3. The recombinant DNA is introduced into cells, where it replicates. 4. Cells containing the desired DNA are identified. Restriction endonucleases are valuable tools for genetic engineers. When different samples of DNA are digested with the same sticky end–generating restriction endonuclease, all the fragments have identical sticky ends. If the fragments are mixed together, the sticky ends can find their complements and re-form base pairs. The discontinuities in the sugar–phosphate backbone can then be mended by a DNA ligase (an enzyme that forms new phosphodiester bonds between adjacent nucleotide residues). These cutting-and-pasting reactions allow a segment of DNA to be excised from a chromosome and inserted into a carrier DNA molecule, such as a circular plasmid, that has been cut by the same restriction enzyme. DNA ligase seals the breaks in the nucleotide strands, leaving an unbroken double-stranded recombinant DNA molecule consisting of the plasmid with a foreign DNA insert (Fig. 3-19). Plasmids are small, circular DNA molecules present in many bacterial cells. A single cell may contain multiple copies of a plasmid, which replicates independently of the bacterial chromosome and usually does not contain genes essential for the host’s normal activities. However, plasmids often do carry genes for specialized functions, such as resistance to certain antibiotics (these genes often encode proteins that inactivate the antibiotics). An antibiotic resistance gene allows the selection of cells that harbor the plasmid: Only cells that contain the plasmid can survive in the presence of the antibiotic. Growing large quantities of plasmid-laden cells is one way to produce large amounts of the foreign DNA insert. (It can be removed later by treating the plasmid with the same restriction enzyme used to insert the foreign DNA.) A piece of DNA that is amplified in this way is said to be cloned. The plasmid that contains the foreign DNA is called a cloning vector. Note that a clone is simply an identical copy of an original. The term is used to refer either to a gene that has been amplified, as described above, or to a cell or organism that is genetically identical to its parent. An example of a cloning vector is shown in Figure 3-20. This plasmid contains a gene (called ampR) for resistance to the antibiotic ampicillin and a gene (called lacZ) encoding the enzyme -galactosidase, which catalyzes the hydrolysis of certain galactose derivatives. The lacZ gene has been engineered to contain several restriction sites, any one of which can be used as an insertion point for a piece of foreign DNA with compatible sticky ends. Interrupting the lacZ gene with foreign DNA prevents the synthesis of the -galactosidase protein. cut Plasmid anneal ligate Recombinant DNA Foreign DNA Figure 3-19 Production of a recombinant DNA molecule. A small circular plasmid and a sample of DNA are cut with the same restriction enzyme, generating complementary sticky ends, so that the fragment of foreign DNA can be ligated into the plasmid. See Animated Figure. Construction of a recombinant DNA molecule. EcoRI SacI KpnI AvaI XmaI SmaI BamHI XbaI SalI AccI HincII PstI SphI HindIII ampR lacZ Figure 3-20 Map of a cloning vector. This 2743-bp circular DNA molecule, called pGEM-3Z, has a gene for resistance to ampicillin so that bacterial cells containing the plasmid can be selected by their ability to grow in the presence of the antibiotic. The plasmid also has a site comprising recognition sequences for 14 different restriction enzymes. Inserting a foreign DNA segment at this site interrupts the lacZ gene, which encodes the enzyme b-galactosidase. Colonies of bacterial cells harboring the intact plasmid can be detected when their b-galactosidase cleaves a galactose derivative that generates a blue dye: HOCH2 HO H HOCH2 Cl O H OH H H OH Galactose Br O H N H HO H Cl O H OH H H OH OH Br HO ⩽ N H H 5-Bromo-4-chloro3-indole Colonies of bacterial cells in which a foreign DNA insert has interrupted the lacZ gene are unable to cleave the galactose derivative and therefore do not turn blue (Fig. 3-21). This method for identifying (screening) bacterial cells containing plasmids with the insert is known as blue-white screening. A single white colony can then be removed from the culture plate and grown in order to harvest its recombinant DNA for sequencing or some other purpose. Bacterial cells that lack the plasmid entirely would also be white. However, these cells are eliminated by including ampicillin in the culture medium, which kills cells that don’t contain the ampR gene. A large number of cloning vectors have been developed to accommodate different sizes of DNA inserts (Table 3-6). Vectors also differ by the cell type in which the vector can grow (for example, bacterial, fungal, insect, or mammalian cells) and in the strategy used to select cells containing the foreign DNA. Cloned genes yield valuable products If a gene that has been isolated and cloned in a host cell is also expressed (transcribed and translated into protein), it may affect the metabolism of that cell. The functions of some gene products have been assessed in this way. Sometimes a specific combination of vector and host cell are chosen so that large quantities of the gene product can be isolated from the cultured cells or from the medium in which they grow. This is an economical method for producing certain proteins that are difficult to obtain directly from human tissues (Table 3-7). After a gene has been isolated and cloned, it can be specifically altered in order to alter the amino acid sequence of the encoded protein. Site-directed mutagenesis (also called in vitro mutagenesis) mimics the natural process of evolution and allows predictions about the structural and functional roles of particular amino acids in a protein to be rigorously tested in the laboratory. One technique for site-directed mutagenesis is a variation of PCR in which the primers are oligonucleotides whose sequences are identical to a portion of the gene of interest except for one or a few bases corresponding to the codon(s) to be altered. The primers can hybridize to the Figure 3-21 Culture dish used in blue-white screening. Blue colonies arise from cells whose plasmids have an intact b-galactosidase gene. White colonies arise from cells whose plasmids contain an insert that interrupts the b-galactosidase gene. [Courtesy S. Kopczak and D. P. Snustad, University of Minnesota.] TABLE 3-6 Types of Cloning Vectors Vector Plasmid Cosmid Bacterial artificial chromosome Yeast artificial chromosome Size of DNA Insert (kb) <20 40–45 40–400 400–1000 78 Ch 3 From Genes to Proteins TABLE 3-7 Some Recombinant Protein Products Protein Purpose Insulin Growth hormone Erythropoietin Treat insulin-dependent diabetes Treat certain growth disorders in children Stimulate production of red blood cells; useful in kidney dialysis Treat hemophilia b and other bleeding disorders Promote clot lysis following myocardial infarction or stroke Promote white blood cell production after bone marrow transplant Coagulation factors IX and X Tissue plasminogen activator Colony stimulating factor wild-type (naturally occurring) gene sequence if there are only a few mismatched bases. Multiple rounds of DNA polymerase–catalyzed primer extension yield many copies of the gene with the desired sequence (Fig. 3-22). The mutagenized gene can then be inserted into a vector for expression in a host cell. Figure 3-22 Site-directed mutagenesis. See Animated Figure. Site-directed mutagenesis. Double-stranded plasmid containing gene to be mutagenized thermal denaturation to separate strands G TC TCCTG C A A C GA GGA T anneal mismatched primers G C A T A C TA C T G G A G G A C T T A C primer extension additional cycles of PCR C G A T T Many copies of double-stranded DNA with mutagenized gene C GA TGA G C T T A C A G A T G AT G A C C T C C T G Substituted base pair 79 Tools and Techniques: Manipulating DNA Genetically modified organisms have practical applications Introducing a foreign gene into a single host cell via a cloning vector alters the genetic makeup of that cell and all its descendants. Producing a transgenic organism whose cells all contain the foreign gene is more difficult. In mammals, the DNA must be injected into fertilized eggs, which are then implanted in a foster mother. Some of the resulting embryos’ cells (possibly including their reproductive cells) will contain the foreign gene. When the animals mature, they must be bred in order to yield offspring whose cells are all transgenic. Transgenic plants are produced by introducing recombinant DNA into a few cells, which can often develop into an entire plant whose cells all contain the foreign DNA. Desirable traits such as resistance to insect pests have been introduced into several important crop species. For example, about two-thirds of the U.S. corn (maize) crop is genetically modified to produce a protein that is toxic to plant-eating insects. So-called Bt corn plants have been engineered to express an insecticidal toxin from the bacterium Bacillus thuringiensis. However, the extensive acreage devoted to the transgenic corn (also known as a genetically modified, or GM, food) raises some concerns: It increases the selective pressure on insects to evolve resistance to the toxin, and it increases the likelihood that the toxin gene will be transferred to other plant species, with disastrous effects on the insects that feed on them. Transgenic plants have also been engineered for better nutrition. For example, researchers have developed strains of rice with foreign genes, from other plant species, that encode enzymes necessary to synthesize b-carotene (an orange pigment that is the precursor of vitamin A) and a gene for the iron-storage protein ferritin (Fig. 3-23). The genetically modified rice strains are intended to help alleviate vitamin A deficiencies (which afflict some 400 million people) and iron deficiencies (an estimated 30% of the world’s population suffers from iron deficiency). Using techniques similar to those for producing transgenic animals, researchers can produce animals in which a particular gene has been inactivated. These “gene knockouts” can serve as animal models for human diseases. For example, mice lacking the cystic fibrosis gene have been used to study the development of the disease and to test potential treatments. Gene therapy can cure some human diseases The goal of gene therapy is to introduce a functional gene into an individual in order to compensate for an existing malfunctioning gene. The first successful gene therapy trials treated children with severe combined immunodeficiency (SCID), a normally fatal condition caused by the absence of the enzyme adenosine deaminase or the absence of a particular receptor protein. Bone marrow cells were removed from each patient, genetically modified, and infused into the patient, where they differentiated into functional immune system cells. The “correct” gene sequence was generated by site-directed mutagenesis and then introduced into the bone marrow cells via a viral vector. Vectors derived from viruses are effective because viruses have evolved to be efficient gene-delivery systems for mammalian cells. Some of the diseases treated by gene therapy are listed in Table 3-8. Despite some 30 years of research effort, significant challenges remain. For example, the viral TABLE 3-8 Some Hereditary Diseases Treated by Gene Therapy Disease Symptoms Adrenoleukodystrophy Hemophilia Leber’s congenital amaurosis Severe combined immunodeficiency (SCID) -Thalassemia Wiskott-Aldrich syndrome Neurodegeneration Bleeding Blindness Immunodeficiency Anemia Immunodeficiency Figure 3-23 Genetically modified rice. The plant has been genetically engineered to produce more b-carotene, the precursor of vitamin A. As a result, the rice grains are yellow rather than white. [© phloen/Alamy.] 80 Ch 3 From Genes to Proteins vectors used in gene therapy may elicit an immune response that destroys the vector before it can deliver its cargo or, as occurred in one case, kills the patient. Moreover, viruses tend to insert themselves into the host cell’s chromosomes somewhat at random, which may interrupt the function of other genes. This has led to several cases of leukemia (uncontrolled growth of an immune system cell) in gene therapy recipients. An additional limitation of the current gene therapy approaches is that most viruses can accommodate a foreign gene no larger than about 8 kb, whereas some human genes have a total length of hundreds of kb (the CF gene is 250 kb). Some genes also require additional regulatory sequences, at some distance from the coding segments, in order for the gene to be expressed in a tissue-specific and developmentally appropriate manner. Whereas a person’s immune system can be “re-seeded” using a few altered cells that continue to multiply, other types of tissues do not readily take up foreign DNA or cannot proliferate to an extent that would permit the introduced gene to compensate for the defective gene. CONCEPT REVIEW • Summarize the steps of dideoxy DNA sequencing. How does pyrosequencing differ from the dideoxy approach? • How do researchers reconstruct long DNA sequences? • Describe the polymerase chain reaction (PCR). Why is PCR a step in DNA fingerprinting? • Why are restriction endonucleases useful for constructing recombinant DNA? • How can cells containing recombinant DNA molecules be identified by selection or screening? • Describe how a cloned gene can be used to produce an altered protein, to generate a transgenic organism, or to cure a disease. SUMMARY 3-1 DNA Is the Genetic Material • The genetic material in virtually all organisms consists of DNA, a polymer of nucleotides. A nucleotide contains a purine or pyrimidine base linked to a ribose group (in RNA) or a deoxyribose group (in DNA) that also bears one or more phosphate groups. • DNA contains two antiparallel helical strands of nucleotides linked by phosphodiester bonds. Each base pairs with a complementary base in the opposite strand: A with T and G with C. The structure of RNA, which is single-stranded and contains U rather than T, is more variable. • Nucleic acid structures are stabilized primarily by stacking interactions between bases. The separated strands of DNA can reanneal. 3-2 Genes Encode Proteins • The central dogma summarizes how the sequence of nucleotides in DNA is transcribed into RNA, which is then translated into protein according to the genetic code. • The sequence of nucleotides in a segment of DNA can reveal mutations that cause disease. 3-3 Genomics • Genomes contain variable amounts of repetitive and other forms of noncoding DNA in addition to genes, which are identified by their sequence characteristics or similarity to other genes. • Genetic variations can be linked to human diseases even when specific disease genes have not been identified. 3-4 Tools and Techniques: Manipulating DNA • The sequence of nucleotides in a segment of DNA is commonly determined by the dideoxy method, in which labeled complementary copies of a DNA strand are synthesized. The presence of dideoxynucleotides, which cannot support further synthesis, generates a set of fragments of different lengths that are separated and analyzed to deduce the sequence of the original template strand. • Large amounts of a particular DNA segment can be obtained by the polymerase chain reaction, in which a DNA polymerase makes complementary copies of a selected segment of DNA. • DNA molecules can be reproducibly fragmented by the action of restriction enzymes, which cleave DNA at specific sequences. • DNA fragments can be joined to each other to generate recombinant DNA molecules that are then cloned (copied) in host cells. • A cloned gene can be manipulated so that large amounts of its protein product can be produced. The sequence of the gene may first be altered by site-directed mutagenesis. • Genes can be introduced into foreign hosts to produce transgenic organisms. In gene therapy, a normal gene is introduced in order to cure a genetic disease. 81 GLOSSARY TERMS chromosome nucleic acid nucleotide DNA (deoxyribonucleic acid) bacteriophage polynucleotide base purine pyrimidine RNA (ribonucleic acid) nucleoside deoxynucleotide vitamin phosphodiester bond residue 59 end 39 end base pair sugar–phosphate backbone antiparallel major groove minor groove bp kb oligonucleotide polymerase nuclease exonuclease endonuclease A-DNA B-DNA stacking interactions BIOINFORMATICS melting temperature (Tm) denaturation renaturation anneal probe replication gene gene expression transcription translation genome coding strand noncoding strand messenger RNA (mRNA) ribosome ribosomal RNA (rRNA) transfer RNA (tRNA) codon genetic code mutation cystic fibrosis (CF) DNA marker splicing diploid haploid transposable element moderately repetitive DNA highly repetitive DNA open reading frame (ORF) homologous genes orphan gene genome map PROJECT 2 horizontal gene transfer genomics single-nucleotide polymorphism (SNP) genome-wide association study (GWAS) dideoxy DNA sequencing ddNTP primer dNTP electrophoresis pyrosequencing polymerase chain reaction (PCR) quantitative PCR (qPCR) restriction endonuclease palindrome DNA fingerprinting allele locus sticky ends blunt ends restriction digest restriction fragment recombinant DNA technology DNA ligase plasmid selection cloning vector clone blue-white screening site-directed mutagenesis transgenic organism gene therapy Learn to locate nucleotide sequences, examine their encoded information, and identify similarities to other sequences. DATABASES FOR THE STORAGE AND “MINING” OF GENOME SEQUENCES PROBLEMS 3-1 DNA Is the Genetic Material 1. The identification of DNA as the genetic material began with Griffith’s “transformation” experiment conducted in 1928. Griffith worked with Pneumococcus, an encapsulated bacterium that causes pneumonia. Wild-type Pneumococcus forms smooth colonies when plated on agar and causes death when injected into mice. A mutant Pneumococcus lacking the enzymes needed to synthesize the polysaccharide capsule (required for virulence) forms rough colonies when plated on agar and does not cause death when injected into mice. Griffith found that heat-treated wild-type Pneumococcus did not cause death when injected into the mice because the heat treatment destroyed the polysaccharide capsule. However, if Griffith mixed heat-treated wild-type Pneumococcus and the mutant unencapsulated Pneumococcus together and injected this mixture, the mice died. Even more surprisingly, upon autopsy, Griffith found live, encapsulated Pneumococcus bacteria in the mouse tissue. Griffith concluded that the mutant Pneumococcus had been “transformed” into disease-causing Pneumococcus, but he could not explain how this occurred. Using your current knowledge of how DNA works, explain how the mutant Pneumococcus became transformed. 2. In 1944, Avery, MacLeod, and McCarty set out to identify the chemical agent capable of transforming mutant unencapsulated Pneumococcus to the deadly encapsulated form. They isolated a viscous substance with the chemical and physical properties of DNA and showed that it was capable of transformation. Transformation could still occur if proteases (enzymes that degrade proteins) or ribonucleases (enzymes that degrade RNA) were added prior to the experiment. What did these treatments tell the investigators about the molecular identity of the transforming factor? 82 3. In 1952, Alfred Hershey and Martha Chase carried out experiments using bacteriophages, which consist of nucleic acid enclosed by a protein capsid (coat). Hershey and Chase first labeled the bacteriophages with the radioactive isotopes 35S and 32P. Because proteins contain sulfur but not phosphorus, and DNA contains phosphorus but not sulfur, each type of molecule was separately labeled. The radiolabeled bacteriophages were allowed to infect the bacteria, and then the preparation was treated to separate the empty capsids (ghosts) from the bacterial cells. The ghosts were found to contain most of the 35S label, whereas 30% of the 32P was found in the new bacteriophages produced by the infected cells. What does this experiment reveal about the roles of bacteriophage DNA and protein? 4. In February 1953 (two months before Watson and Crick published their paper describing DNA as a double helix), Linus Pauling and Robert Corey published a paper in which they proposed that DNA adopts a triple-helical structure. In their model, the three chains are tightly packed together. The phosphate groups reside on the inside of the triple helix, while the nitrogenous bases are located on the outside. They proposed that the triple helix was stabilized by hydrogen bonds between the interior phosphate groups. What are the flaws in this model? 5. In some organisms, DNA is modified by addition of methyl groups. Draw the structure of 5-methylcytosine. 6. In certain pathogenic bacteria, the methylation of certain adenines in DNA is required in order for the bacteria to cause disease. (a) Draw the structure of N 6-methyladenine. (b) Why might scientists be interested in the fact that other bacteria produce an N 6-DNA methyltransferase (an enzyme that catalyzes the transfer of methyl groups to adenine)? 7. Certain strains of E. coli incorporate the nitrogenous base shown here into nucleotides. For which base is this one a substitute? O Cl 15. A well-studied bacteriophage has 97,004 bases in its complete genome. (a) There are 24,182 G residues in the genome. Calculate the number of C, A, and T residues. (b) Why does GenBank report a total of 48,502 bases for this bacteriophage genome? 16. The complete genome of a virus contains 1578 T residues, 1180 G residues, 1609 A residues, and 1132 C residues. What can you conclude about the structure of the viral genome, given this information? 17. Identify the base pair highlighted in blue in Figure 3-7. 18. The adenine derivative hypoxanthine can base pair with cytosine, adenine, and uracil. Show the structures of these base pairs. N N O 8. An E. coli culture is grown in the presence of the base shown in Problem 7. A control culture is grown in the absence of this modified base. Compare the masses of the DNA isolated from E. coli in these two cultures. 9. Describe the chemical difference between uracil and thymine. 10. (a) What kind of linkage joins the two nucleotides in the dinucleotides NAD and FAD (see Figure 3-3)? (b) How do the adenosine groups in FAD and CoA differ? N H Hypoxanthine 19. Explain whether the following statement is true or false: Because a G:C base pair is stabilized by three hydrogen bonds, whereas an A:T base pair is stabilized by only two hydrogen bonds, GC-rich DNA is harder to melt than AT-rich DNA. 20. Hydrogen bonding does not make a significantly large contribution to the overall stability of the DNA molecule. Explain. 21. How can the hydrophobic effect (Section 2-2) explain why DNA adopts a helical structure? 22. (a) Would you expect proteins to bind to the major groove or the minor groove of DNA? Explain. (b) Eukaryotic DNA is packaged with histones, small proteins with a high lysine and arginine content. Why do histones have a high affinity for DNA? NH N H NH O O HN CH C O NH CH C CH2 CH2 CH2 CH2 CH2 CH2 CH2 NH NH⩽ 3 C Lysine NH⩽ 2 NH2 Arginine 11. Draw a CA (ribo)dinucleotide and label the phosphodiester bond. How would the structure differ if it were DNA? 23. Draw melting curves that would be obtained from the DNA of Dictyostelium discoideum and Streptomyces albus (see Table 3-2). 12. Many cellular signaling pathways involve the conversion of ATP to cyclic AMP, in which a single phosphate group is esterified to both C39 and C59. Draw the structure of cyclic AMP. 24. Using Table 3-2 as a guide, estimate the melting temperature of the DNA from an organism whose genome contains equal amounts of all four nucleotides. 13. Do Chargaff ’s rules hold true for RNA? Explain why or why not. 25. What might you find in comparing the GC content of DNA from Thermus aquaticus or Pyrococcus furiosus and DNA from bacteria in a typical backyard pond? 14. A diploid organism with a 30,000-kb haploid genome contains 19% T residues. Calculate the number of A, C, G, and T residues in the DNA of each cell in this organism. 26. Explain why the melting temperature of a sample of doublehelical DNA increases when the Na1 concentration increases. 83 27. You have a short piece of synthetic RNA that you want to use as a probe to identify a gene in a sample of DNA. The RNA probe has a tendency to hybridize with sequences that are only weakly complementary. Should you increase or decrease the temperature to improve your chances of tagging the correct sequence? 28. In the laboratory technique known as fluorescence in situ hybridization (FISH), a fluorescent oligonucleotide probe is allowed to hybridize with a cell’s chromosomes, which are typically spread on a microscope slide. Explain why the chromosome preparation must be heated before the probe is added to it. 3-2 Genes Encode Proteins 29. Discuss the shortcomings of the following definitions for gene: (a) A gene is the information that determines an inherited characteristic such as flower color. (b) A gene is a segment of DNA that encodes a protein. (c) A gene is a segment of DNA that is transcribed in all cells. 30. The semiconservative nature of DNA replication (as shown in Fig. 3-10) was proposed by Watson and Crick in 1953, but it wasn’t experimentally verified until 1958. Meselson and Stahl grew bacteria on a “heavy nitrogen” source (ammonium chloride containing the isotope 15N) for many generations so that virtually every nitrogen atom in the bacterial DNA was the 15N isotope. This resulted in DNA that was denser than normal. The food source was abruptly switched to one containing only 14N. Bacteria were harvested and the DNA isolated by density gradient centrifugation. (a) What is the density of the DNA of the first-generation daughter DNA molecules? Explain. (b) What is the density of the DNA isolated after two generations? Explain. 31. A segment of the coding strand of a gene is shown below. ACACCATGGTGCATCTGACT (a) Write the sequence of the complementary strand that DNA polymerase would make. (b) Write the sequence of the mRNA that RNA polymerase would make from the gene segment. 32. A portion of a gene is shown below. 59-ATGATTCGCCTCGGGGCTCCCCAGTCGCTGGTGCTG39-TACTAAGCGGAGCCCCGAGGGGTCAGCGACCACGAGCTGACGCTGCTCGTCG-39 CGACTGCGACGAGCAGC-59 The sequence of the mRNA transcribed from this gene has the following sequence: 59-AUGAUUCGCCUCGGGGCUCCCCAGUCGCUGGUGCUGACGCUGCUCGUCG-39 (a) Identify the coding and noncoding strands of the DNA. (b) Explain why only the coding strands of DNA are commonly published in databanks. 33. In the early 1960s, Marshall Nirenberg deciphered the genetic code by designing an experiment in which he synthesized a polynucleotide strand consisting solely of U residues, then added this strand to a test tube containing all of the components needed for protein synthesis. (a) What polypeptide was produced by this “cell-free” system? (b) What polypeptides were produced when poly(A), poly(C), and poly(G) were added to the cell-free system? 34. Har Gobind Korana, using a new polynucleotide synthesis method that he developed, extended Nirenberg’s work by synthesizing polynucleotides with precisely defined sequences. (a) What polypeptide(s) would be produced if a poly(GU) were added to the cell-free system described in Problem 33? (b) Do these results help to decipher the identities of the codons involved? 35. An open reading frame (ORF) is a portion of the genome that potentially codes for a protein. A given nucleotide sequence of mRNA potentially has three different reading frames, only one of which is correct (the selection of the correct ORF will be discussed more fully in Section 22-3). A portion of the gene for a type II human collagen is shown. What are the sequences of amino acids that can potentially be translated from each of the three possible reading frames? AGGTCTTCAGGGAATGCCTGGCGAGAGGGGAGCAGCTGGTATCGCTGGGCCCAAAGGC 36. Refer to Problem 35. Collagen’s amino acid sequence consists of repeating triplets in which every third amino acid is glycine. Does this information assist you in your identification of the correct reading frame? 37. One form of the disease adrenoleukodystrophy (ALD) is caused by the substitution of serine for asparagine in the ALD protein. List the possible single-nucleotide alterations in the DNA of the ALD gene that could cause this genetic disease. 38. In another form of adrenoleukodystrophy, a CGA codon in the ALD gene is converted to a UGA codon. Explain how this mutation affects the ALD protein. 39. A mutation occurs when there is a base change in the DNA sequence. Some base changes do not lead to changes in the amino acid sequence of the resulting protein. Explain why. 40. Is it possible for the same segment of DNA to encode two different proteins? Explain. 41. A portion of the nucleotide sequence from the DNA coding strand of the chick ovalbumin gene is shown. What is the partial amino acid sequence of the encoded protein? CTCAGAGTTCACCATGGGCTCCATCGGTGCAGCAAGCATGGAA— (1104 bp)— TTCTTTGGCAGATGTGTTTCCCCTTAAAAAGAA 42. A type of gene therapy called RNA interference (RNAi) is being investigated to treat Huntington’s disease. This disease is the result of a mutation in the DNA that results in the synthesis of a nervous system protein with an altered amino acid sequence. The mutated protein forms clumps, which cause nervous system defects. To treat this disease, scientists synthesize short sequences of RNA (siRNA, or small interfering RNA) that form base pairs with the mRNA that codes for the mutated protein. (a) Design an siRNA that will interfere with the synthesis of the protein shown in Problem 41. (b) Explain how the addition of the siRNA will prevent the synthesis of the protein. (c) What are the difficulties that must be overcome in order for RNA interference to be an effective technique for treating the disease? 3-3 Genomics 43. The genome of the bacterium Carsonella ruddii contains 159 kb of DNA with 182 ORFs. What can you conclude about the habitat or lifestyle of this bacterium? 84 44. In theory, both strands of DNA can code for proteins; that is, genes can be overlapping. Propose an explanation for why overlapping genes are more commonly observed in prokaryotes than in eukaryotes. 50. Assuming that genes and SNPs are distributed evenly throughout the human genome, estimate how many protein-coding genes are likely to differ between two individuals. 45. For many years, biologists and others have claimed that humans and chimpanzees are 98% identical at the level of DNA. Both the human and chimp genomes, which are roughly the same size, have now been sequenced, and the data reveal approximately 35 million nucleotide differences between the two species. How does this number compare to the original claim? 51. A genome-wide association study was carried out to identify the SNPs located on chromosome 1 that were correlated with an intestinal disease. The location of three genes on chromosome 1 (between positions 6.3 3 107 and 6.6 3 107) is shown in the figure below. A 21og10 P-value of 7 or greater is assumed to be associated with the disease. (a) What are the locations on the chromosome that show the strongest correlation with the disease? (b) Which genes contain SNPs associated with the disease and which do not? 47. A partial sequence of a newly discovered bacteriophage is shown below. (a) Identify the longest open reading frame (ORF). (b) Assuming that the ORF has been correctly identified, where is the most likely start site? TATGGGATGGCTGAGTACAGCACGTTGAATGAGGCGATGGCCGCTGGTGATG 48. The bacteriophage described in Problem 47 contains 59 kb and 105 ORFs. None of the ORFs code for tRNAs. How does the bacteriophage replicate its DNA and synthesize the structural proteins necessary to replicate itself? 49. If each person’s genome contains a SNP every 300 nucleotides or so, how many SNPs are in that person’s genome? gene A −log10 P-value 46. When genomes of various organisms were sequenced, biologists expected that the DNA content (the C-value) would be positively correlated with organismal complexity. But no such correlation has been demonstrated. In fact, some amphibians have DNA content that is similar to that found in humans. The C-value paradox is the term that refers to this puzzling lack of correlation between DNA content and organismal complexity. What questions do biologists need to ask as they attempt to solve the paradox? gene C gene B 14 12 10 8 6 4 2 0 67,400,000 67,300,000 67,500,000 52. Use the information in the figure below to determine the chromosomal locations for the SNPs that are most closely associated with a colon disease. 22.5 20.0 −log10 P-value 17.5 15.0 12.5 10.0 7.5 5.0 2.5 0 Chr1 Chr13 Chr2 Chr14 Chr3 Chr15 Chr4 Chr5 Chr16 Chr6 Chr17 Chr7 Chr18 Chr8 Chr19 Chr9 Chr20 Chr10 Chr21 Chr11 Chr22 Chr12 Chrx 85 3-4 Tools and Techniques: Manipulating DNA 53. Design a 10-bp primer that could be used to determine the following sequence of DNA: bottom of the gel, whereas the larger fragments are found near the top of the gel.) Control EcoRI 59-AGTCGATCCCTGATCGTACGCTACGGTAACGT-39 54. The following fragments were produced by using the “shotgun” method. Show how the fragments should be aligned to determine the sequence of the DNA. ACCGTGTTTCCGACCG ATTGTTCCCACAGACCG CGGCGAAGCATTGTTCC TTGTTCCCACAGACCGTG 55. The primer used in sequencing a cloned DNA segment often includes the recognition sequence for a restriction endonuclease. Explain. 56. Heat-stable DNA polymerases are sometimes used for dideoxy sequencing at high temperatures, especially when the template DNA has a high content of G 1 C. Explain. 57. Could you perform PCR with an ordinary DNA polymerase, that is, one that is destroyed by high temperatures? What modifications would you make in the PCR protocol? 58. Examine the sequence of the protein in Solution 41. Assume that the corresponding DNA sequence is not known. Using the amino acid sequence as a guide, design a pair of nine-base deoxynucleotide primers that could be used for PCR amplification of the protein-coding portion of the gene. (Hint: DNA polymerase can extend a primer only from its 39 end.) How many different pairs of primers could you choose from? 59. Which restriction enzymes in Table 3-5 generate sticky ends? Blunt ends? 60. A sample of the DNA segment shown below is treated with the restriction enzyme MspI. Next, the sample is incubated with an exonuclease that acts only on single-stranded polynucleotides. What mononucleotides will be present in the reaction mixture? TGCTTAGCCGGA ACGA AC GAATCGGCCTTGCT PstI EcoRI + PstI 20 kb 15 kb 11 kb 9 kb 8 kb 7 kb 3 kb 2 kb Use the results of the agarose gel electrophoresis separation (above) to construct a restriction map for the sample of DNA. 63. The search for a gene sometimes starts with a DNA library, which is a set of cloned DNA fragments representing all the sequences in an organism’s genome. If you were to construct a human DNA library to search for novel genes, would you choose to clone the DNA fragments in plasmids or in yeast artificial chromosomes? Explain. 64. A researcher trying to identify the gene for a known protein might begin by looking closely at the protein’s sequence in order to design a single-stranded oligonucleotide probe that will hybridize with the DNA of the gene. Why would the researcher focus on a segment of the protein containing a methionine (Met) or tryptophan (Trp) residue (see Table 3-3)? 65. A group of investigators is interested in studying the gp41 protein from the human immunodeficiency virus (HIV). In order to do this, they use site-directed mutagenesis to synthesis a series of truncated proteins. A partial sequence of the gp41 protein is shown. Design an 18 bp “mismatched primer” (see Fig. 3-22) that could be used to synthesize a truncated protein that would terminate after the Leu residue at position 700. 700 61. Which is more likely to be called a “rare cutter”: a restriction enzyme with a four-base recognition sequence or a restriction enzyme with an eight-base recognition sequence? . . . WYIKLFIMIVGGLVGLRIVFAVLSIVNRVRGGYSP . . . 62. Restriction enzymes are used to construct restriction maps of DNA. These are diagrams of specific DNA molecules that show the sites where the restriction enzymes cleave the DNA. To construct a restriction map, purified samples of the DNA are treated with restriction enzymes, either alone or in combination, and then the reaction products are separated by agarose gel electrophoresis, a technique that separates the DNA fragments based on size. (The smaller fragments travel more quickly and are found near the 66. Refer to the DNA sequence shown in Problem 41. (a) Using the coding sequence for this gene, design two 18-bp PCR primers that could be used to amplify this gene. (b) Suppose you wanted to add EcoRI restriction sites to each end of the gene. How would you modify the sequences of the PCR primers you designed in (a) to amplify the gene? 86 SELECTED READINGS Collins, F. S., Cystic fibrosis: molecular biology and therapeutic implications, Science 256, 774–779 (1992). [Describes the discovery of the CFTR gene and its protein product.] Dickerson, R. E., DNA structure from A to Z, Methods Enzymol. 211, 67–111 (1992). [Describes the various crystallographic forms of DNA.] Fischer, A., and Cavazzana-Calvo, M., Gene therapy of inherited diseases, Lancet 371, 2044–2047 (2008). [Discusses strategies, progress, and ongoing challenges in human gene therapy.] Galperin, M. Y., and Fernández-Suárez, X. M., The 2012 Nucleic Acids Research Database Issue and the online Molecular Biology Database Collection, Nucleic Acids Res. 40, D1–D8 (2012). [Briefly describes some new additions to a set of 1380 databases related to molecular biology. Also available at http://nar.oxfordjournals.org/.] International Human Genome Sequencing Consortium, Initial sequencing and analysis of the human genome, Nature 409, 860–921 (2001) and Venter, J. C., et al., The sequence of the human genome, Science 291, 1304–1351 (2001). [These and other papers in the same issues of Nature and Science describe the data that constitute the draft sequence of the human genome and discuss how this information can be used in understanding biological function, evolution, and human health.] Lander, E. S., Initial impact of the sequencing of the human genome, Nature 470, 187–197 (2011). [Reviews genome structure and provides insights into its variation and links to human diseases.] Thieffry, D., Forty years under the central dogma, Trends Biochem. Sci. 23, 312–316 (1998). [Traces the origins, acceptance, and shortcomings of the idea that nucleic acids contain biological information.] chapter 4 PROTEIN STRUCTURE HOW do faulty protein structures lead to disorders such as Alzheimer’s disease? The effects of this neurodegenerative condition are clearly visible when comparing scans of the brain of an Alzheimer’s disease patient (right) and a normal brain (left). In this chapter, we’ll examine protein structure, beginning with the amino acids that form polypeptide chains. We’ll also see how proteins are stabilized by weak noncovalent forces, including the hydrophobic effect. When some protein structures are disrupted, the molecules tend to aggregate, eventually impairing cellular functions and producing the symptoms of disorders such as Alzheimer’s disease. [Dr. Robert Friedland/Photo Researchers, Inc.] THIS CHAPTER IN CONTEXT Do You Remember? Part 1 Foundations Part 2 Molecular Structure and Function 4 Protein Structure Cells contain four major types of biological molecules and three major types of polymers (Section 1-2). Noncovalent forces, including hydrogen bonds, ionic interactions, and van der Waals forces, act on biological molecules (Section 2-1). The hydrophobic effect, which is driven by entropy, excludes nonpolar substances from water (Section 2-2). An acid’s pK value describes its tendency to ionize (Section 2-3). Part 3 Metabolism The biological information encoded by a sequence of DNA is transcribed to RNA and then translated into the amino acid sequence of a protein (Section 3-2). Part 4 Genetic Information 87 88 Ch 4 Protein Structure Proteins are the workhorses of the cell. They provide structural stability and motors for movement; they form the molecular machinery for harvesting free energy and using it to carry out other metabolic activities; they participate in the expression of genetic information; and they mediate communication between the cell and its environment. In subsequent chapters, we will describe in more detail these proteindriven phenomena, but for now we will focus on protein structure. Proteins come in many shapes and sizes (Fig. 4-1). The essence of their biological function is their interaction with other molecules, including other proteins. We will look first at the amino acid components of proteins. Next comes a discussion of how the protein backbone folds and how the backbone plus all the side chains assume a unique three-dimensional shape. The Tools and Techniques section of this chapter examines some of the procedures for purifying and sequencing proteins and determining their three-dimensional shapes. Plastocyanin (poplar) Shuttles electrons as part of the apparatus for converting light energy to chemical energy (more in Section 16-2) DNA polymerase (E. coli Klenow fragment) Synthesizes a new DNA chain using an existing DNA strand as a template (more in Section 20-2) Insulin (pig) Released from the pancreas to signal the availability of the metabolic fuel glucose (more in Section 19-2) Maltoporin (E. coli) Permits sugars to cross the bacterial cell membrane (more in Section 9-2) Figure 4-1 A gallery of protein structures. These space- filling models are all shown at approximately the same scale. In proteins that consist of more than one chain of amino acids, the chains are shaded differently. [Structure of insulin (pdb 1ZNI) determined by M. G. W. Turkenburg, J. L. Whittingham, G. G. Dodson, E. J. Dodson, B. Xiao, and G. A. Bentley; structure of maltoporin (pdb 1MPM) Phosphoglycerate kinase (yeast) Catalyzes one of the central reactions in metabolism (more in Section 13-1) determined by R. Dutzler and T. Schirmer; structure of phosphoglycerate kinase (pdb 3PGK) determined by P. J. Shaw, N. P. Walker, and H. C. Watson; structure of DNA polymerase (pdb 1KFS) determined by C. A. Brautigan and T. A. Steitz; and structure of plastocyanin (pdb 1PND) determined by B. A. Fields, J. M. Guss, and H. C. Freeman.] 89 4-1 Proteins Are Chains of Amino Acids A protein is a biological molecule that consists of one or more polypeptides, which are chains of polymerized amino acids. A cell may contain dozens of different amino acids, but only 20 of these—called the “standard” amino acids—are commonly found in proteins. As introduced in Section 1-2, an amino acid is a small molecule – containing an amino group (ONH1 3) and a carboxylate group (OCOO) as well as a side chain of variable structure, called an R group: COO⩾ H C KEY CONCEPTS • The 20 amino acids differ in the chemical characteristics of their R groups. • Amino acids are linked by peptide bonds to form a polypeptide. • A protein’s structure may be described at four levels, from primary to quaternary. R NH⩽ 3 Note that at physiological pH, the carboxyl group is unprotonated and the amino group is protonated, so an isolated amino acid bears both a negative and a positive charge. The 20 amino acids have different chemical properties The identities of the R groups distinguish the 20 standard amino acids. The R groups can be classified by their overall chemical characteristics as hydrophobic, polar, or charged, as shown in Figure 4-2, which also includes the one- and three-letter codes for each amino acid. These compounds are formally called a-amino acids because the amino and carboxylate (acid) groups are both attached to a central carbon atom known as the a carbon (abbreviated Ca). Nineteen of the twenty standard amino acids are asymmetric, or chiral, molecules. Their chirality, or handedness (from the Greek cheir, “hand”), results from the asymmetry of the alpha carbon. The four different substituents of Ca can be arranged in two ways. For alanine, a small amino acid with a methyl R group, the possibilities are ⩾ H3N⩽ Cα CH3 H Does Chirality Matter? The importance of chirality in biological systems was brought home in the 1960s when pregnant women with morning sickness were given the sedative thalidomide, which was a mixture of right- and left-handed forms. The active form of the drug has the structure shown below. H COO H Cα O NH⩽ 3 CH3 You can use a simple model-building kit to satisfy yourself that the two structures are not identical. They are nonsuperimposable mirror images, like right and left hands. The amino acids found in proteins all have the form on the left. For historical reasons, these are designated L amino acids (from the Greek levo, “left”). Their mirror images, which rarely occur in proteins, are the D amino acids (from dextro, “right”). Molecules related by mirror symmetry are physically indistinguishable and are usually present in equal amounts in synthetic preparations. However, the two forms behave differently in biological systems (Box 4-A). It is advisable to become familiar with the structures of the standard amino acids, since their side chains ultimately help determine the three-dimensional shape of the protein as well as its chemical reactivity. The Hydrophobic Amino Acids As their name implies, the hydrophobic amino acids have essentially nonpolar side chains that interact very weakly or not at all with water. The aliphatic (hydrocarbonlike) side chains of alanine (Ala), valine (Val), leucine (Leu), isoleucine (Ile), and phenylalanine (Phe) obviously fit into this group. Although methionine (Met, with an S atom) and tryptophan (Trp, with an NH group) include atoms that can form hydrogen bonds, the bulk of their side chains is nonpolar. Proline (Pro) is unique among the amino acids because its aliphatic side chain is also covalently linked to its amino group. O N ⩾ COO BIOCHEMISTRY NOTE BOX 4-A H O C N O Thalidomide Tragically, its mirror image, which was also present, caused severe birth defects, including abnormally short or absent limbs. Although the mechanisms of action of the two forms of thalidomide are not well understood, different responses to the two forms can be rationalized. An organism’s ability to distinguish chiral molecules results from the handedness of its molecular constituents. For example, proteins contain all L amino acids, and polynucleotides coil in a right-handed helix (see Fig. 3-4). The lessons learned from thalidomide have made drugs more costly to develop and test but have also made them safer. Question: Which of the 20 amino acids is not chiral? In proteins, the hydrophobic amino acids are almost always located in the interior of the molecule, among other hydrophobic groups, where they do not interact with water. And because they lack reactive functional groups, the hydrophobic side chains do not directly participate in mediating chemical reactions. 90 Ch 4 Protein Structure The Polar Amino Acids The side chains of the polar amino acids can interact with water because they contain hydrogen-bonding groups. Serine (Ser), threonine (Thr), and tyrosine (Tyr) have hydroxyl groups; cysteine (Cys) has a thiol group; and asparagine (Asn) and glutamine (Gln) have amide groups. All these amino acids, along with histidine (His, which bears a polar imidazole ring), can be found on the solvent-exposed surface of a protein, although they also occur in the protein interior, provided that their hydrogen-bonding Hydrophobic amino acids COO⩾ CH3 COO⩾ H C CH3 H NH⩽ 3 ⩾ COO H C CH C NH⩽ 3 Alanine (Ala, A) H CH3 ⩾ CH3 H C H CH3 C CH CH2 H C CH2 CH2 N H COO⩾ CH2 H C CH2 H2N⩽ CH2 S CH3 NH⩽ 3 Isoleucine (Ile, I) CH2 Tryptophan (Trp, W) COO CH3 C NH⩽ 3 ⩾ COO Leucine (Leu, L) CH2 Phenylalanine (Phe, F) NH⩽ 3 CH3 COO⩾ NH⩽ 3 Valine (Val, V) CH2 CH NH⩽ 3 COO⩾ Methionine (Met, M) Proline (Pro, P) Polar amino acids COO⩾ CH3 COO⩾ H C CH2 OH H NH⩽ 3 H C OH H C NH⩽ 3 COO⩾ O CH2 C NH2 H NH⩽ 3 C CH2 CH2 CH2 OH H C NH2 H C Glutamine (Gln, Q) CH2 SH COO⩾ N CH2 NH⩽ 3 C NH⩽ 3 Cysteine (Cys, C) COO⩾ O NH⩽ 3 Asparagine (Asn, N) COO⩾ NH⩽ 3 Tyrosine (Tyr, Y) Threonine (Thr, T) Serine (Ser, S) COO⩾ CH C COO⩾ H N H C H NH⩽ 3 Histidine (His, H) Glycine (Gly, G) Charged amino acids COO⩾ H C COO⩾ O CH2 C O NH⩽ 3 Aspartate (Asp, D) ⩾ H C COO⩾ O CH2 CH2 C O ⩾ H NH⩽ 3 C COO⩾ CH2 CH2 CH2 NH⩽ 3 Glutamate (Glu, E) Figure 4-2 Structures and abbreviations of the 20 standard amino acids. The amino acids can be classified according to the chemical properties of their R groups as hydrophobic, polar, or charged. The side chain (R group) of each amino acid is shaded. The three-letter code is usually the first three letters of the amino acid’s name. The one-letter code is derived as follows: If only one amino acid begins with a particular letter, that letter is used: C 5 cysteine, H 5 histidine, I 5 isoleucine, M 5 methionine, S 5 serine, and V 5 valine. If more than one amino acid begins with a particular letter, the letter is assigned to the most abundant amino acid: A 5 alanine, G 5 glycine, L 5 leucine, P 5 proline, and T 5 threonine. Most of the others are CH2 NH⩽ 3 H C NH2 CH2 CH2 CH2 NH C NH⩽ 2 NH⩽ 3 Lysine (Lys, K) Arginine (Arg, R) phonetically suggestive: D 5 aspartate (“asparDate”), F 5 phenylalanine (“Fenylalanine”), N 5 asparagine (“asparagiNe”), R 5 arginine (“aRginine”), W 5 tryptophan (“tWyptophan”), and Y 5 tyrosine (“tYrosine”). The rest are assigned as follows: E 5 glutamate (near D, aspartate), K 5 lysine, and Q 5 glutamine (near N, asparagine). The carbon atoms of amino acids are sometimes assigned Greek letters, beginning with Ca, the carbon to which the R group is attached. Thus, glutamate has a g-carboxylate group, and lysine has an ε-amino group. ? Identify the functional groups in each amino acid. Refer to Table 1-1. requirements are satisfied by their proximity to other hydrogen bond donor or acceptor groups. Glycine (Gly), whose side chain consists of only an H atom, cannot form hydrogen bonds but is included with the polar amino acids because it is neither hydrophobic nor charged. Depending on the presence of nearby groups that increase their polarity, some of the polar side chains can ionize at physiological pH values. For example, the neutral (basic) form of His can accept a proton to form an imidazolium ion (an acid): COO⩾ HC N CH2 HC N H NH⩽ 3 COO⩾ H⩽ H⩽ NH⩽ CH2 N H NH⩽ 3 Base Acid As we will see, the ability of His to act as an acid or a base gives it great versatility in catalyzing chemical reactions. Similarly, the thiol group of Cys can be deprotonated, yielding a thiolate anion: COO⩾ HC COO⩾ H⩽ CH2 SH HC H⩽ NH⩽ 3 CH2 S⩾ NH⩽ 3 Occasionally, cysteine’s thiol group undergoes oxidation with another thiol group, such as another Cys side chain, to form a disulfide bond: COO⩾ HC COO⩾ CH2 S S CH2 NH⩽ 3 CH NH⩽ 3 Disulfide bond In rare cases, the very weakly acidic hydroxyl groups of Ser, Thr, and Tyr ionize to yield hydroxide groups that can act as strong bases in chemical reactions. The Charged Amino Acids Four amino acids have side chains that are virtually always charged under physiological conditions. Aspartate (Asp) and glutamate (Glu), which bear carboxylate groups, are negatively charged. Lysine (Lys) and arginine (Arg) are positively charged. These side chains are usually located on the protein’s surface, where their charged groups can be surrounded by water molecules or interact with other polar or ionic substances. Although it is convenient to view amino acids merely as the building blocks of proteins, many amino acids play key roles in regulating physiological processes (Box 4-B). Peptide bonds link amino acids in proteins The polymerization of amino acids to form a polypeptide chain involves the condensation of the carboxylate group of one amino acid with the amino 91 Proteins Are Chains of Amino Acids 92 Ch 4 Protein Structure BOX 4-B BIOCHEMISTRY NOTE Monosodium Glutamate A number of amino acids and compounds derived from them function as signaling molecules in the nervous system (we will look at some of these in more detail in Section 18-2). Among the amino acids with signaling activity is glutamate, which most often operates as an excitatory signal and is necessary for learning and memory. Because glutamate is abundant in dietary proteins and because the human body can manufacture it, glutamate deficiency is rare. But is there any danger in eating too much glutamate? Glutamate binds to receptors on the tongue that register the taste of umami—one of the five human tastes, along with sweet, salty, sour, and bitter. By itself, the umami taste is not particularly pleasing, but when combined with other tastes, it imparts a sense of savoriness and induces salivation. For this reason, glutamate in the form of monosodium glutamate (MSG) is sometimes added to processed foods as a flavor enhancer. For example, a low-salt food item can be made more appealing by adding MSG to it. According to some popular accounts, “Chinese restaurant syndrome” can be attributed to the consumption of excess MSG added to prepared foods or present in soy sauce (MSG is also naturally present in many other foods, including cheese and tomatoes). The symptoms of the syndrome reportedly include muscle tingling, headache, and drowsiness—all of which could potentially reflect the role of glutamate in the nervous system. However, a definitive link between MSG intake and neurological symptoms has not been demonstrated in scientific studies and therefore remains mostly anecdotal. Question: Draw the structure of MSG. Which group is ionized in order to pair with a single sodium ion? What is the structure of magnesium glutamate? Would it have the same physiological effects as MSG? group of another (a condensation reaction is one in which a water molecule is eliminated): R1 ⩽ H3N C H H O ⩽ C R2 ⩽ H N ⩾ O C H H O C O⩾ H2O ⩽ H3N R1 O R2 C C H C N H H O C O⩾ Peptide bond The resulting amide bond linking the two amino acids is called a peptide bond. The remaining portions of the amino acids are called amino acid residues. In a cell, peptide bond formation is carried out in several steps involving the ribosome and additional RNA and protein factors (Section 22-3). Peptide bonds can be broken, or hydrolyzed, by the action of exo- or endopeptidases (enzymes that act from the end or the middle of the chain, respectively). By convention, a chain of amino acid residues linked by peptide bonds is written or drawn so that the residue with a free amino group is on the left (this end of the TABLE 4-1 93 Proteins Are Chains of Amino Acids pK Values of Ionizable Groups in Polypeptides Groupa pK C-terminus COOH 3.5 O Asp CH2 C OH 3.9 O Glu CH2 His CH2 C OH 4.1 NH⩽ CH2 6.0 N H Cys CH2 N-terminus NH⩽ 3 Tyr CH2 Lys CH2 SH 8.4 9.0 OH CH2 CH2 10.5 CH2 NH⩽ 3 10.5 NH2 Arg a CH2 CH2 CH2 NH C NH⩽ 2 12.5 The ionizable proton is indicated in red. polypeptide is called the N-terminus) and the residue with a free carboxylate group is on the right (this end is called the C-terminus): R1 N-terminus ⩽ H 3N O R2 O R3 O R4 O CH C N CH C N CH C N CH C H Residue 1 H Residue 2 O⩾ H Residue 3 Residue 4 Note that, except for the two terminal groups, the charged amino and carboxylate groups of each amino acid are eliminated in forming peptide bonds. The electrostatic properties of the polypeptide therefore depend primarily on the identities of the side chains (R groups) that project out from the polypeptide backbone. The pK values of all the charged and ionizable groups in amino acids are given in Table 4-1 (recall from Section 2-3 that a pK value is a measure of a group’s tendency to ionize). Thus, it is possible to calculate the net charge of a protein at a given pH (see Sample Calculation 4-1). At best, this value is only an estimate, since the side chains of polymerized amino acids do not behave as they do in free amino acids. This is because of the electronic effects of the peptide bond and other functional groups that may be brought into proximity when the polypeptide chain folds into a three-dimensional shape. The chemical properties of a side chain’s immediate neighbors, its microenvironment, may alter its polarity, thereby altering its tendency to lose or accept a proton. Nevertheless, the chemical and physical properties of proteins depend on their constituent amino acids, so proteins exhibit different behaviors under given laboratory conditions. These differences can be exploited to purify a protein, that is, to isolate it from a mixture containing other molecules (see Section 4-5). C-terminus SAMPLE CALCULATION 4-1 PROBLEM Estimate the net charge of the polypeptide chain below at physiological pH (7.4) and at pH 5.0. Ala–Arg–Val–His–Asp–Gln SOLUTION The polypeptide contains the following ionizable groups, whose pK values are listed in Table 4-1: the N-terminus (pK 5 9.0), Arg (pK 5 12.5), His (pK 5 6.0), Asp (pK 5 3.9), and the C-terminus (pK 5 3.5). At pH 7.4, the groups whose pK values are less than 7.4 are mostly deprotonated, and the groups with pK values greater than 7.4 are mostly protonated. The polypeptide therefore has a net charge of 0: Group Charge N-terminus 11 Arg 11 His 0 Asp 21 C-terminus 21 net charge 0 At pH 5.0, His is likely to be protonated, giving the polypeptide a net charge of 11: Group Charge N-terminus 11 Arg 11 His 11 Asp 21 C-terminus 21 net charge 11 PRACTICE PROBLEMS 1. Estimate the net charge of a Glu–Tyr dipeptide at pH 6.0. 2. Estimate the net charge of an Asp–Asp–Asp tripeptide at pH 7.0. 3. Estimate the net charge of a His–Lys–Glu tripeptide at pH 8.0. Most polypeptides contain between 100 and 1000 amino acid residues, although some contain thousands of amino acids (Table 4-2). Polypeptides smaller than about 40 residues are often called oligopeptides (oligo is Greek for “few”) or just peptides. Since there are 20 different amino acids that can be polymerized to form polypeptides, even peptides of similar size can differ dramatically from each other, depending on their complement of amino acids. TABLE 4-2 Composition of Some Proteins Protein Insulin (bovine) Rubredoxin (Pyrococcus) Myoglobin (human) Phosphorylase kinase (yeast) Hemoglobin (human) Reverse transcriptase (HIV) Nitrite reductase (Alcaligenes) C-reactive protein (human) Pyruvate decarboxylase (yeast) Immunoglobulin (mouse) Ribulose bisphosphate carboxylase (spinach) Glutamine synthetase (Salmonella) Carbamoyl phosphate synthetase (E. coli) Number of Amino Acid Residues Number of Polypeptide Chains Molecular Mass (D) 51 53 153 416 574 986 1029 1030 1112 1316 5048 5628 5820 2 1 1 1 4 2 3 5 2 4 16 12 8 5733 5878 17,053 44,552 61,972 114,097 111,027 115,160 121,600 145,228 567,960 621,600 637,020 The potential for sequence variation is enormous. For a modest-sized polypeptide of 100 residues, there are 20100 or 1.27 3 10130 possible amino acid sequences. This number is clearly unattainable in nature, since there are only about 1079 atoms in the universe, but it illustrates the tremendous structural variability of proteins. Unraveling the amino acid sequence of a protein may be relatively straightforward if its gene has been sequenced (see Section 3-3). In this case, it is just a matter of reading successive sets of three nucleotides in the DNA as a sequence of amino acids in the protein. However, this exercise may not be accurate if the gene’s mRNA is spliced before being translated or if the protein is hydrolyzed or otherwise covalently modified immediately after it is synthesized. And, of course, nucleic acid sequencing is of no use if the protein’s gene has not been identified. The alternative is to use a technique such as mass spectrometry to directly determine the protein’s amino acid sequence (Section 4-5). The amino acid sequence is the first level of protein structure The sequence of amino acids in a polypeptide is called the protein’s primary structure. There are as many as four levels of structure in a protein (Fig. 4-3). Under physiological conditions, a polypeptide very seldom assumes a linear extended conformation but instead folds up to form a more compact shape, usually consisting of Primary structure The sequence of amino acid residues –Glu–Ser–Phe–Gly–Asp– Secondary structure The localized conformation of the polypeptide backbone Tertiary structure The three-dimensional structure of an entire polypeptide, including all its side chains Quaternary structure The spatial arrangement of polypeptide chains in a protein with multiple subunits Figure 4-3 Levels of protein structure in hemoglobin. [Structure of human hemoglobin (pdb 2HHB) determined by G. Fermi and M. F. Perutz.] 95 Proteins Are Chains of Amino Acids 96 Ch 4 Protein Structure several layers. The local folding arrangement of the polypeptide backbone (exclusive of the side chains) is known as secondary structure. The complete three-dimensional conformation of the polypeptide, including its backbone atoms and all its side chains, is the polypeptide’s tertiary structure. In a protein that consists of more than one polypeptide chain, the quaternary structure refers to the spatial arrangement of all the chains. In the following sections we will consider the second, third, and fourth levels of protein structure. CONCEPT REVIEW • Draw the structures and give the one- and three-letter abbreviations for the 20 standard amino acids. • Divide the 20 amino acids into groups that are hydrophobic, polar, and charged. • Which polar amino acids are sometimes charged? • Describe the four levels of protein structure. 4-2 Secondary Structure: The Conformation of the Peptide Group KEY CONCEPTS • The polypeptide backbone has limited conformational flexibility. • The a helix and b sheet are common secondary structures characterized by hydrogen bonding between backbone groups. In the peptide bond that links successive amino acids in a polypeptide chain, the electrons are somewhat delocalized so that the peptide bond has two resonance forms: O⩾ O C C ⩽ N N H H Due to this partial (about 40%) double-bond character, there is no rotation around the CON bond. In a polypeptide backbone, the repeating NOCaOC units of the amino acid residues can therefore be considered to be a series of planar peptide groups (where each plane contains the atoms involved in the peptide bond): H N O Cα C H N H Cα C O N O Cα C N H Cα C O Here the H atom and R group attached to Ca are not shown. The polypeptide backbone can still rotate around the NOCa and CaOC bonds, although rotation is somewhat limited. For example, a sharp bend at Ca could bring the carbonyl oxygens of neighboring residues too close: Here the atoms are color-coded: C gray, O red, N blue, and H

white. As the resonance structures on the opposite page indicate, the groups involved in the peptide bond are strongly polar, with a tendency to form hydrogen bonds. The backbone amide groups are hydrogen bond donors, and the carbonyl oxygens are hydrogen bond acceptors. Under physiological conditions, the polypeptide chain folds so that it
can satisfy as many of these hydrogen-bonding requirements as possible. At the same time, the polypeptide backbone must adopt a conformation (a secondary structure) that minimizes steric strain. In addition, side chains must be positioned in a way that minimizes their steric interference. Two kinds of secondary structure commonly found in
proteins meet these criteria: the a helix and the b sheet. 97 Secondary Structure: The Conformation of the Peptide Group The a helix exhibits a twisted backbone conformation The a helix was first identified through model-building studies carried out by Linus Pauling. In this type of secondary structure, the polypeptide backbone twists in a right-
handed helix (the DNA helix is also right-handed; see Section 3-1 for an explanation). There are 3.6 residues per turn of the helix, and for every turn, the helix rises 5.4 Å along its axis. In the a helix, the carbonyl oxygen of each residue forms a hydrogen bond with the backbone NH group four residues ahead. The backbone hydrogen-bonding
tendencies are thereby met, except for the four residues at each end of the helix (Fig. 4-4). Most a helices are about 12 residues long. Like the DNA helix, whose side chains fill the helix interior (see Figure 3-4b), the a helix is solid—the atoms of the polypeptide backbone are in van der Waals contact. However, in the a helix, the side chains extend
outward from the helix (Fig. 4-5). The b sheet contains multiple polypeptide strands Pauling, along with Robert Corey, also built models of the b sheet. This type of secondary structure consists of aligned strands of polypeptide whose hydrogenbonding requirements are met by bonding between neighboring strands. The strands of a b sheet can be
arranged in two ways (Fig. 4-6): In a parallel b sheet, neighboring chains run in the same direction; in an antiparallel b sheet, neighboring chains run in opposite directions. Each residue forms two hydrogen bonds with a neighboring strand, so all hydrogen-bonding requirements are met, except in the first and last strands of the sheet. Figure 4-4 The
a helix. In this conformation, the polypeptide backbone twists in a right-handed fashion so that hydrogen bonds (dashed lines) form between CPO and NOH groups four residues farther along. Atoms are color-coded: Ca light gray, carbonyl C dark gray, O red, N blue, side chain purple, H white. [Based on a drawing by Irving Geis.] Figure. The a helix. ?
See Animated How many amino acid residues are shown here? How many hydrogen bonds? See Guided Exploration. Secondary structures in proteins. See Guided Exploration. Stable helices in proteins: the a helix. (a) (b) Figure 4-5 An helix from myoglobin. In (a) a ball-and-stick model and (b) a space-filling model of residues 100–118 of myoglobin,
the backbone atoms are green and the side chains are gray. [Structure of sperm whale myoglobin (pdb 1MBD) determined by S. E. V. Phillips.] 98 N N C C O O C O C O C C H N H N H N C C O C O C N H N H C C C O N H C C C O O C O C N H C C C O H N C N H N H C H N C C C C N Antiparallel Parallel See Guided Exploration. Hydrogen bonding
in b sheets. O O H N C C N H O C C O C C C O H N C C N H C C C C O N H H N C C O C C C N H C C O O C O O C C N H C C O H N C N H N H N H C C N H C O H N C C C C O H N C O C C O H N C N C C N H C C O O N H C N C C N H O N Figure 4-6 b sheets. In a parallel b sheet and an antiparallel b sheet, the polypeptide backbone is
extended. In both types of b sheet, hydrogen bonds form between the amide and carbonyl groups of adjacent strands. The H and R attached to Ca are not shown. Note that the strands are not necessarily separate polypeptides but may be segments of a single chain that loops back on itself. See Animated Figure. b sheets. ? How many amino acid
residues are shown in each chain? A single b sheet may contain from 2 to more than 12 polypeptide strands, with an average of 6 strands, and each strand has an average length of 6 residues. In a b sheet, the amino acid side chains extend from both faces (Fig. 4-7). Proteins also contain irregular secondary structure a Helices and b sheets are
classified as regular secondary structures, because their component residues exhibit backbone conformations that are the same from one residue to the next. In fact, these elements of secondary structure are easily recognized in the (a) (b) Figure 4-7 Side view of two parallel strands of a b sheet. In (a) a ball-and-stick model and (b) a space-filling
model of a b sheet from carboxypeptidase A, the backbone atoms are green. The amino acid side chains (gray) point alternately to each side of the b sheet. [Structure of carboxypeptidase (pdb 3CPA) determined by W. N. Lipscomb.] three-dimensional structures of a huge variety of proteins, regardless of their amino acid composition. Of course,
depending on the identities of the side chains and other groups that might be present, a helices and b sheets may be slightly distorted from their ideal conformations. For example, the final turn of some a helices becomes “stretched out” (longer and thinner than the rest of the helix). In every protein, elements of secondary structure (individual
helices or strands in a sheet) are linked together by polypeptide loops of various sizes. A loop may be a relatively simple hairpin turn, as in the connection of two antiparallel b strands (which are shown below as flat arrows; left). Or it may be quite long, especially if it joins successive strands in a parallel b sheet (right): 99 Tertiary Structure and
Protein Stability Usually, the loops that link b strands or a helices consist of residues with irregular secondary structure; that is, the polypeptide does not adopt a defined secondary structure in which successive residues have the same backbone conformation. Note that “irregular” does not mean “disordered”: The peptide backbone almost always
adopts a single, unique conformation. Most proteins contain a combination of regular and irregular secondary structure. On average, 31% of residues are in a helices, 28% are in b sheets, and most of the remainder are in irregular loops of different sizes. CONCEPT REVIEW • • • • • Which amino acid atoms constitute the polypeptide backbone? What
limits the rotation of the backbone? Summarize the structural features of a helices and b sheets. Distinguish regular and irregular secondary structure. Why does virtually every protein contain loops? 4-3 Tertiary Structure and Protein Stability The three-dimensional shape of a protein, known as its tertiary structure, includes its regular and irregular
secondary structure (that is, the overall folding of its peptide backbone) as well as the spatial arrangement of all its side chains. In a fully folded protein under physiological conditions, virtually every atom has a designated place. One of the most powerful techniques for probing the atomic structures of macromolecules, including proteins, is X-ray
crystallography (Section 4-5). The structure of myoglobin, the first protein structure to be determined by X-ray crystallography, came to light in 1958 through the efforts of John Kendrew, who painstakingly determined the conformation of every backbone and side-chain group. Kendrew’s results—coming just a few years after Watson and Crick had
published their elegant model of DNA—were a bit of a disappointment. The myoglobin structure lacked the simplicity and symmetry of a molecule such as DNA and was more irregular and complex than expected (myoglobin structure is examined in detail in Section 5-1). The structure of another well-studied protein, the enzyme triose phosphate
isomerase, is shown in Figure 4-8a. This enzyme is a globular protein (fibrous proteins, in contrast, are usually highly elongated; examples of these are presented in Section 5-2). The proteins shown in Figure 4-1 are all globular proteins. The tertiary structure of KEY CONCEPTS • A folded polypeptide assumes a shape with a hydrophilic surface and a
hydrophobic core. • Protein folding and protein stabilization depend on noncovalent forces. • Some proteins can adopt more than one stable conformation. 100 Ch 4 Protein Structure triose phosphate isomerase can be simplified by showing just the peptide backbone (Fig. 4-8b). Alternatively, the structure can be represented by a ribbon that passes
through Ca of each residue (Fig. 4-8c). This rendering makes it easier to identify elements of secondary structure. Commonly used systems for analyzing protein structures are based on the presence of secondary structural elements. For example, four classes are recognized by the CATH system (the name refers to a hierarchy of organizational levels:
Class, Architecture, Topology, and Homology). Proteins may contain mostly a structure, mostly b structure, a combination of a and b, or very few secondary structural elements. Examples of each class are shown in Figure 4-9. Structural data— including the three-dimensional coordinates for each atom—for tens of thousands of proteins and other
macromolecules are available in online databases. The use of software for visualizing and manipulating such structures provides valuable insight into molecular structure and function (see Bioinformatics Project 3, Visualizing Three-Dimensional Protein Structures). Proteins have hydrophobic cores (a) Globular proteins typically contain at least two
layers of secondary structure. This means that the protein has definite surface and core regions. On the protein’s surface, some backbone and side-chain groups are exposed to the solvent; in the core, (b) (b) (a) (c) Figure 4-8 Triose phosphate isomerase. (a) Space-filling model. All (c) (d) atoms (except H) are shown (C gray, O red, N blue). (b)
Polypeptide backbone. The trace connects the a carbons of successive amino acid residues. (c) Ribbon diagram. The ribbon represents the overall conformation of the backbone. (b) gB-Crystallin, an all-b protein. [Structure (pdb 1GCS) determined by S. Najmudin, P. Lindley, C. Slingsby, O. Bateman, D. Myles, S. Kumaraswamy, and I. Glover.] (c)
Flavodoxin, an a/b protein. [Structure (pdb 1CZN) determined by W. W. Smith, K. A. Pattridge, C. L. Luschinsky, and M. L. Ludwig.] (d) Tachystatin, a protein with little secondary structure. [Structure [Structure (pdb 1YPI) determined by T. Alber, E. Lolis, and G. A. Petsko.] (pdb 1CIX) determined by N. Fujitani, S. Kawabata, T. Osaki, Y. Kumaki, M.
Demura, K. Nitta, and K. Kawano.] Figure 4-9 Classes of protein structure. In each protein, a helices are colored red, and b strands are colored yellow. (a) Growth hormone, an all-a protein. [Structure (pdb 1HGU) determined by L. Chantalat, N. Jones, F. Korber, J. Navaza, and A. G. Pavlovsky.] these groups are sequestered from the solvent. In other
words, the protein comprises a hydrophilic surface and a hydrophobic core. A polypeptide segment that has folded into a single structural unit with a hydrophobic core is often called a domain. Some small proteins consist of a single domain. Larger proteins may contain several domains, which may be structurally similar or dissimilar (Fig. 4-10). The
core of a small hydrophobic protein (or a domain) is typically rich in regular secondary structure. This is because the formation of a helices and b sheets, which are internally hydrogen-bonded, minimizes the hydrophilicity of the polar backbone groups. Irregular secondary structures (loops) are more often found on the surface of the protein (or
domain), where the polar backbone groups can form hydrogen bonds to water molecules. The requirement for a hydrophobic core and a hydrophilic surface also places constraints on amino acid sequence. The location of a particular side chain in a protein’s tertiary structure is related to its hydrophobicity: The greater a residue’s hydrophobicity, the
more likely it is to be located in the protein interior. In the protein interior, side chains pack together, leaving very little empty space or space that could be occupied by a water molecule. Table 4-3 lists two scales for assessing the hydrophobicity of amino acid side chains. Such information is useful for predicting the locations of amino acid residues
within proteins. For example, highly hydrophobic residues such as Phe and Met are almost always buried. Polar side chains, like hydrogen-bonding backbone groups, can participate in hydrogen bonding in the protein interior, which helps “neutralize” their polarity and allows them to be buried in a nonpolar environment. When a charged residue
occurs in the protein interior, it is almost always located near another residue with the opposite charge, so the two groups can interact electrostatically to form an ion pair. By color-coding the amino acid residues of myoglobin according to their hydrophobicity, it is easy to see that hydrophobic side chains cluster in the interior while hydrophilic side
chains predominate on the surface (Fig. 4-11). TABLE 4-3 Residue Phe Met Ile Leu Val Cys Trp Ala Thr Gly Ser Pro Tyr His Gln Asn Glu Lys Asp Arg Figure 4-10 A two-domain protein. In this model of glyceraldehyde-3phosphate dehydrogenase, the small domain is red, and the large domain is green. [Structure (pdb 1GPD) determined by D. Moras, K.
W. Olsen, M. N. Sabesan, M. Buehner, G. C. Ford, and M. G. Rossmann.] ? Which of the proteins in Figure 4-9 consists of two domains? Hydrophobicity Scales Scale Aa Scale Bb 2.8 1.9 4.5 3.8 4.2 2.5 20.9 1.8 20.7 20.4 20.8 21.6 21.3 23.2 23.5 23.5 23.5 23.9 23.5 24.5 3.7 3.4 3.1 2.8 2.6 2.0 1.9 1.6 1.2 1.0 0.6 20.2 20.7 23.0 24.1 24.8 28.2 28.8 29.2
212.3 a Scale A is from Kyte, J., and Doolittle, R.F., J. Mol. Biol. 157, 105–132 (1982). b 101 Scale B is from Engelman, D.M., Steitz, T.A., and Goldman, A., Annu. Rev. Biophys. Chem. 15, 321–353 (1986). (a) (b) Figure 4-11 Hydrophobic and hydrophilic residues in myoglobin. (a) All hydrophobic side chains (Ala, Ile, Leu, Met, Phe, Pro, Trp, and Val) are
colored green. These residues are located mostly in the protein interior. (b) All polar and charged side chains are colored red. These are located primarily on the protein surface. 102 Ch 4 Protein Structure (a) Folded (b) Unfolded Unfavorable solvation Figure 4-12 The hydrophobic effect in protein folding. (a) In a folded protein, hydrophobic regions
(represented by blue segments of the polypeptide chain) are sequestered in the protein interior. (b) Unfolding the protein exposes these segments to water. This arrangement is energetically unfavorable, because the presence of the hydrophobic groups interrupts the hydrogen-bonded network of water molecules. Protein structures are stabilized
mainly by the hydrophobic effect Surprisingly, the fully folded conformation of a protein is only marginally more stable than its unfolded form. The difference in thermodynamic stability amounts to about 0.4 kJ ? mol–1 per amino acid, or about 40 kJ ? mol–1 for a 100-residue polypeptide. This is equivalent to the amount of free energy required to break
two hydrogen bonds (about 20 kJ ? mol21 each). This quantity seems incredibly small, considering the number of potential interactions among all of a protein’s backbone and side-chain atoms. Nevertheless, most proteins do fold into a unique and stable three-dimensional arrangement of atoms. The largest force governing protein structure is the
hydrophobic effect (introduced in Section 2-2), which causes nonpolar groups to aggregate in order to minimize their contact with water. The hydrophobic effect is driven by the increase in entropy of the solvent water molecules, which would otherwise have to order themselves around each hydrophobic group. As we have seen, hydrophobic side
chains are located predominantly in the interior of a protein. This arrangement stabilizes the folded polypeptide backbone, since unfolding it or extending it would expose the hydrophobic side chains to the solvent (Fig. 4-12). Hydrogen bonding by itself is not a major determinant of protein stability, because in an unfolded protein, polar groups could
just as easily form energetically equivalent hydrogen bonds with water molecules. Instead, hydrogen bonding may help the protein finetune a folded conformation that is already largely stabilized by the hydrophobic effect. Cross-links help stabilize proteins Many folded polypeptides appear to be held in place by cross-links of various kinds, the most
common being ion pairs, disulfide bonds, and inorganic ions such as zinc. Do these cross-links help stabilize the protein? An ion pair can form between oppositely charged side chains or the N- and C-terminal groups (Fig. 4-13). Although the resulting electrostatic interaction is strong, it does not contribute much to protein stability. This is because the
favorable free energy of the electrostatic interaction is offset by the loss of entropy when the side chains become fixed in the ion pair. For a buried ion pair, there is the additional energetic cost of desolvating the charged groups in order for them to enter the hydrophobic core. + 103 Tertiary Structure and Protein Stability − Lys Glu (a) Asp − + Arg
(b) Figure 4-13 Examples of ion pairs in myoglobin. (a) The ε-amino group of Lys 77 interacts with the carboxylate group of Glu 18. (b) The carboxylate group of Asp 60 interacts with Arg 45. The atoms are color-coded: C gray, N blue, and O red. Note that these intramolecular interactions occur between side chains that are near each other in the
protein’s tertiary structure but are far apart in the primary structure. ? Identify residues that could form two other types of ion pairs. Disulfide bonds (shown in Section 4.1) can form within and between polypeptide chains. Experiments show that even when the Cys residues of certain proteins are chemically blocked, the proteins may still fold and
function normally. This suggests that disulfide bonds are not essential for stabilizing these proteins. In fact, disulfides are rare in intracellular proteins, since the cytoplasm is a reducing environment. They are more plentiful in proteins that are secreted to an extracellular (oxidizing) environment (Fig. 4-14). Here, the bonds may help prevent protein
unfolding under relatively harsh extracellular conditions. Domains containing cross-links called zinc fingers are common in DNAbinding proteins. These structures consist of 20–60 residues with one or two Zn21 ions. The Zn21 ions are tetrahedrally coordinated by the side chains of Cys and/or His and sometimes Asp or Glu (Fig. 4-15). Protein domains
this size are (a) (b) 21 Figure 4-15 Zinc fingers. (a) A zinc finger with one Zn (purple sphere) coordinated by two Cys and two His residues, from Xenopus transcription factor IIIA. (b) A zinc finger with two Zn21 coordinated by six Cys residues, from the yeast transcription factor GAL4. [Structure of transcription factor IIIA (pdb 1TF6) determined by R.
T. Nolte, R. M. Conlin, S. C. Harrison, and R. S. Brown; structure of GAL4 (pdb 1D66) determined by R. Marmorstein and S. Harrison.] Figure 4-14 Disulfide bonds in lysozyme, an extracellular protein. This 129-residue enzyme from hen egg white contains eight Cys residues (yellow), which form four disulfide bonds that link different sites on the
polypeptide backbone. [Structure (pdb 1E8L) determined by H. Schwalbe, S. B. Grimshaw, A. Spencer, M. Buck, J. Boyd, C. M. Dobson, C. Redfield, and L. J. Smith.] 104 Ch 4 Protein Structure too small to assume a stable tertiary structure without a metal ion cross-link. Zinc is an ideal ion for stabilizing proteins: It can interact with ligands (S, N, or
O) provided by several amino acids, and it has only one oxidation state (unlike Cu or Fe ions, which readily undergo oxidation–reduction reactions under cellular conditions). Protein folding begins with the formation of secondary structures Figure 4-16 Model of protein folding. In this hypothetical example, the polypeptide first forms elements of
secondary structure (a helices and b sheets). These coalesce into a single globular mass, and small adjustments generate the final stable tertiary structure. [After Goldberg, M. E., Trends Biochem. Sci. 10, 389 (1985).] The crowded nature of the cell interior (see Fig. 2-7) demands that proteins and other macromolecules assume compact shapes. In the
cell, a newly synthesized polypeptide begins to fold as soon as it emerges from the ribosome, so part of the polypeptide may adopt its mature tertiary structure before the entire chain has been synthesized. It is difficult to monitor this process in the cell, so studies of protein folding in vitro usually use full-length polypeptides that have been chemically
unfolded (denatured) and then allowed to refold (renature). In the laboratory, proteins can be denatured by adding highly soluble substances such as salts or urea (NH2OCOONH2). Large amounts of these solutes interfere with the structure of the solvent water, thereby attenuating the hydrophobic effect and causing the protein to unfold. When the
solutes are removed, the proteins renature. Protein renaturation experiments demonstrate that protein folding is not a random process. That is, the protein does not just happen upon its most stable tertiary structure (its native structure) by trial and error but approaches this conformation through one or a few alternate pathways. During this process,
small elements of secondary structure form first, then these coalesce under the influence of the hydrophobic effect to produce a mass with a hydrophobic core. Finally, small rearrangements yield the native tertiary structure (Fig. 4-16). All the information required for a protein to fold is contained in its amino acid sequence. Unfortunately, there are
no completely reliable methods for predicting how a polypeptide chain will fold. In fact, it is difficult to determine whether a given amino acid sequence will form an a helix, b sheet, or irregular secondary structure. This presents a formidable obstacle to assigning three-dimensional structures—and possible functions—to the burgeoning number of
proteins identified through genome sequencing (see Section 3-3). In the laboratory, certain small proteins can be repeatedly denatured and renatured, but in the cell, protein folding is more complicated and may require the assistance of other proteins. Some of these are known as molecular chaperones and are described in more detail in Section 22-4.
Proteins that do not fold properly— often due to genetic mutations that substitute one amino acid for another—can lead to a variety of diseases (Box 4-C). The pathway to full functionality may require additional steps beyond polypeptide folding. Some proteins contain several polypeptide chains, which must fold individually before assembling. In
addition, many proteins undergo processing before reaching their mature forms. Depending on the protein, this might mean removal of some amino acid residues or the covalent attachment of another group such as a lipid, carbohydrate, or phosphate group (Fig. 4-17). The attached groups usually have a discrete biological function and may also help
stabilize the folded conformation of the protein. Finally, some proteins become functional only after associating with metal ions or a specific organic molecule. Some proteins can adopt more than one conformation. A protein’s native structure is not necessarily rigid and inflexible. In fact, some minor movement, primarily the result of bending and
stretching of individual bonds, is required for most proteins to carry out their biological functions. In addition, some proteins—as many as 50% in eukaryotes—may include highly flexible extended segments that are rich in hydrophilic amino acids. These intrinsically unstructured proteins or protein segments might be particularly well suited for
regulatory roles, in which one protein must interact with several others. BOX 4-C CLINICAL CONNECTION Protein Misfolding and Disease HOW do faulty protein structures lead to disorders such as Alzheimer’s disease? At the start of the chapter, we posed a question about how faulty protein structures cause disease. The answer lies in the cell’s
failure to deal with misfolded proteins. Normally, the chaperones that help new proteins to fold can also help misfolded proteins to refold. If the protein cannot be salvaged in this way, it is usually degraded to its component amino acids. Newly synthesized polypeptide misfolding aggregation folding Misfolded structure degradation Native structure
refolding The operation of this quality control system explains what happens in the most common form of cystic fibrosis (see Box 3-A): A mutated form of the CFTR protein folds incorrectly and therefore never reaches its intended cellular destination. Other diseases result when misfolded proteins are not immediately refolded or degraded but instead
aggregate, often as long insoluble fibers. Although the fibers can occur throughout the body, they appear to be deadliest when they occur in the brain. Among the human diseases characterized by such fibrous deposits are Alzheimer’s disease, Parkinson’s disease, and the transmissible spongiform encephalopathies, which lead to neurological
abnormalities and loss of neurons (nerve cells). The aggregated proteins, a different type in each disease, are commonly called amyloid deposits (a name originally referring to their starchlike appearance). Alzheimer’s disease, the most common neurodegenerative disease, is accompanied by both intracellular “tangles” and extracellular “plaques.”
Amyloid deposits (large red areas) surrounded by intracellular tangles (smaller dark shapes). [Courtesy Dennis Selkoe and Marcia Podlisny, Harvard University Medical School.] The fibrous material inside cells is made of a protein called tau, which is involved in the assembly of microtubules, a component of the cytoskeleton (Section 5-2). Tau deposits
also appear in some other neurodegenerative diseases, and tau’s role in Alzheimer’s disease is not clear. The extracellular amyloid material in Alzheimer’s disease consists primarily of a protein called amyloid-b, a 40- or 42-residue fragment of the much larger amyloid precursor protein, which is a membrane protein. Proteases called b-secretase and
g-secretase cleave the precursor protein to generate amyloid-b. Normal brain tissue contains some extracellular amyloid-b, but neither its function nor the function of its precursor protein is completely understood. However, excess amyloid-b is clearly linked to Alzheimer’s disease. The neurodegeneration of Alzheimer’s disease may begin many years
before symptoms such as memory loss appear, and studies of the disease in animal models indicate that both of these signs can be detected before amyloid fibers accumulate in the brain. These and other experimental results support the hypothesis that the early stages of amyloid-b misfolding and aggregation are toxic to neurons and are the ultimate
cause of Alzheimer’s disease. The accumulation of extracellular fibers may actually be a protective mechanism to deal with excessive amyloid-b production. In Parkinson’s disease, neurons in a portion of the brain accumulate fragments of a protein known as a-synuclein. Like amyloid-b, a-synuclein’s function is not fully known, but it appears to play a
role in neurotransmission. a-Synuclein is a small soluble protein (140 amino acid residues) with an extended conformation, part of which appears to form a helices upon binding to other molecules. The intrinsic disorder of the protein may contribute to its propensity to form amyloid deposits, which are characterized by a high content of b secondary
structure. Accumulation of this material is associated with the death of neurons, leading to the typical symptoms of Parkinson’s disease, which include tremor, muscular rigidity, and slow movements. Mutations in the gene for a-synuclein that lead to increased expression of the protein or promote its self-aggregation are responsible for some
hereditary forms of Parkinson’s disease. Mad cow disease is the best known of the transmissible spongiform encephalopathies (TSEs), a group of disorders that also includes scrapie in sheep and Creutzfeldt–Jakob disease in humans. These fatal diseases, which give the brain a spongy appearance, were once thought to be caused by a virus. However,
extensive investigation has revealed that the infectious agent is a protein called a prion. Interestingly, normal human brain tissue contains the same protein, named PrPC (C for cellular), which occurs on neural cell membranes and appears to play a role in normal brain function. The scrapie form of the prion protein, PrPSc, has the same 253–amino
acid sequence as PrPC but includes more b secondary structure. Introduction of PrPSc into cells apparently triggers PrPC, which contains more a-helical structure, to assume the PrPSc conformation and thereby aggregate with it. How is a prion disease transmitted? One route seems to be by ingestion, as illustrated by the incidence of bovine TSE in
cows consuming feed prepared from scrapie-infected sheep. (This feeding practice has been discontinued, but not before a number of people—who presumably consumed the infected beef— developed a variant form of Creutzfeldt–Jakob disease). PrPSc must be absorbed, without being digested, and transported to the central nervous system. Here, it
causes neurodegeneration, possibly through the loss of PrPC as it converts to PrPSc or through the toxic effects of the PrPSc as it accumulates. Despite the lack of sequence or structural similarities among amyloid-b, a-synuclein, and PrPSc, their misfolded forms are all rich in b structure, and this seems to be the key to the formation of amyloid
deposits. Studies of a fungal polypeptide similar to PrPSc show how amyloid formation might occur. In its original state, the protein is mostly a-helical. A segment of polypeptide shifts to an all-b conformation, which allows single molecules to stack on top of each other in parallel to form a triangular fiber stabilized by interchain hydrogen bonding. In
this model, five aggregated polypeptides are shown. The vertical arrow indicates the long axis of the amyloid fiber. Presumably, other amyloid-forming proteins undergo similar conformational changes. Significant aggregation may not occur until the protein concentration reaches some critical threshold (this may explain why the amyloid diseases,
even the TSEs, take years to develop). After a few b structures have assembled, they act to trigger further protein misfolding, and the amyloid fibers rapidly propagate. Once formed, the fibers are resistant to degradation by proteases. Model for amyloid formation. [From Meier, B., et al., Science, 319, 1523–1526 (2008). Reprinted with permission
from AAAS.] spongiform encephalopathy after being inoculated with PrPSc, which causes disease in normal mice. What do these results reveal about the relationship between PrP and TSE? What do they reveal about the cellular function of PrP? 3. Myoglobin is a protein that contains mostly a helices and no b strands (see Fig. 5-1). (a) Would you
expect myoglobin to form amyloid fibers? (b) Under certain laboratory conditions, myoglobin can be induced to form amyloid fibers. What does this suggest about a polypeptide’s ability to adopt b secondary structure? 4. Discuss the role of amino acid side chains in the formation of an amyloid fiber. Questions: 1. The gene for the amyloid precursor
protein is located on chromosome 21. Individuals with Down syndrome (trisomy 21) have three rather than two copies of chromosome 21. Explain why individuals with Down syndrome tend to exhibit Alzheimer’s-like symptoms by middle age. 2. Mice that have been genetically engineered to lack the PrP gene (but are normal in all other ways) do not
develop In some cases, a protein’s conformational flexibility includes two stable alternatives in dynamic equilibrium. A change in cellular conditions, such as pH or oxidation state, or the presence of a binding partner can tip the balance toward one conformation or the other (Fig. 4-18). Because structural biologists have traditionally examined proteins
with fixed, stable shapes, newer approaches to studying protein structure will most likely reveal more about intrinsically unstructured proteins and proteins with more than one conformational variant. (b) (a) (c) NH O HOCH2 H3C (CH2)14 C H NH O S CH2 O CH C O O H OH H NH C CH2 CH C H O⪖ ⪖ H HN C O CH3 O P O NH O CH2 CH C O O
Figure 4-17 Some covalent modifications of proteins. (a) A 16-carbon fatty acid (palmitate, in red) is linked by a thioester bond to a Cys residue. (b) A chain of several carbohydrate units (here only one sugar residue is shown, in red) is linked to the amide N of an Asn side chain. (c) A phosphoryl group (red) is esterified to a Ser side chain. 107
Quaternary Structure (a) (b) Figure 4-18 A protein with two stable conformations. One form of the protein lymphotactin consists of a three-stranded b sheet and an b helix (a). This form interconverts rapidly with an alternate form (b) with an all-b structure. [Structures (a, pdb 1J9O and b, pdb 2JP1) determined by B. F. Volkman.] CONCEPT REVIEW •
How do the surface and core of a protein differ? Which residues are likely to be found in each region? • What forces stabilize protein structure? • How does an unfolded polypeptide reach its native conformation? 4-4 Quaternary Structure Most proteins, especially those with molecular masses greater than 100 kD, consist of more than one polypeptide
chain. The individual chains, called subunits, may all be identical, in which case the protein is known as a homodimer, homotetramer, and so on. If the chains are not all identical, the prefix hetero- is used. The spatial arrangement of these polypeptides is known as the protein’s quaternary structure. The forces that hold subunits together are similar to
those that determine the tertiary structures of the individual polypeptides. The interface (the area of contact) between two subunits is often hydrophobic, with closely packed side chains. Hydrogen bonds, ion pairs, and disulfide bonds dictate the exact geometry of the interacting subunits. Among the most common quaternary structures in proteins
are symmetrical arrangements of two or more identical subunits (Fig. 4-19). Even in proteins with nonidentical subunits, the symmetry is based on groups of subunits. For example, hemoglobin, a heterotetramer with two ab units, can be considered to be a dimer of dimers (see Fig. 4-19c). A few multimeric proteins are known to adopt more than one
possible quaternary structure and apparently shift among the alternatives by dissociating, adopting different tertiary structures, and reassembling. The advantages of multisubunit protein structure are many. For starters, the cell can construct extremely large proteins through the incremental addition of small building blocks (we will see examples of
this in the next chapter). This is clearly an asset for certain structural proteins that—due to their enormous size—cannot be synthesized all in one piece or must be assembled outside the cell. Moreover, the impact of the inevitable errors in transcription and translation can be minimized if the affected polypeptide is small and readily replaced. Finally,
the interaction between subunits in a multisubunit protein affords an opportunity for the subunits to influence each other’s behavior or work cooperatively. The result is a way of regulating function that is not possible in single-subunit proteins or in multisubunit proteins whose subunits each operate independently. In Chapter 5, we will examine the
cooperative behavior of hemoglobin, which has four interacting oxygen-binding sites. CONCEPT REVIEW • What are the advantages of quaternary structure? KEY CONCEPT • Proteins containing more than one polypeptide chain have quaternary structure. 108 Ch 4 Protein Structure Figure 4-19 Some proteins with quaternary structure. The alpha
carbon backbone of each polypeptide is shown. (a) Nitrite reductase, an enzyme with three identical subunits, from Alcaligenes. [Structure (pdb 1AS8) determined by M. E. P Murphy, E. T. Adams, and S. Turley.] (b) E. coli fumarase, a homotetrameric enzyme. [Structure (pdb 1FUQ) determined by T. Weaver and L. Banaszak.] (c) Human hemoglobin,
a heterotetramer with two a subunits (blue) and 2 b subunits (red). [Structure (pdb 2HHB) determined by G. Fermi and M. F. Perutz.] (d) Bacterial methane hydroxylase, whose two halves (right and left in this image) each contain three kinds of subunits. (a) (b) (c) (d) [Structure (pdb 1MMO) determined by A. C. Rosenzweig, C. A. Frederick, S. J.
Lippard, and P. Nordlund.] ? Which of the proteins shown in Fig. 4-1 has quaternary structure? 4-5 Tools and Techniques: Analyzing Protein Structure KEY CONCEPTS • Chromatography is a technique for separating molecules on the basis of size, charge, or specific binding behavior. • The sequence of amino acids in a polypeptide can be determined
by mass spectrometry. • The three-dimensional arrangement of atoms in a protein can be deduced by measuring the diffraction of X-rays or electrons or by analyzing nuclear magnetic resonance. Like nucleic acids (Section 3-4), proteins can be purified and analyzed in the laboratory. In this section, we will examine some commonly used approaches to
isolating proteins and determining their sequence of amino acids and their three-dimensional structures. Chromatography takes advantage of a polypeptide’s unique properties As described in Section 4-1, a protein’s amino acid sequence determines its overall chemical characteristics, including its size, shape, charge, and ability to interact with other
substances. A variety of laboratory techniques have been devised to exploit these features in order to separate proteins from other cellular components. One of the most powerful techniques is chromatography. Originally performed with solvents moving across paper, chromatography now typically uses a column packed with a porous matrix (the
stationary phase) and a buffered solution (the mobile phase) that percolates through the column. Proteins or other solutes pass through the column at different rates, depending on how they interact with the stationary phase. In size-exclusion chromatography (also called gel filtration chromatography), the stationary phase consists of tiny beads with
pores of a characteristic size. If a solution containing proteins of different sizes is applied to the top of the column, the proteins will move through the column as fluid drips out the bottom. Larger proteins will be excluded from the spaces inside the beads and will pass through the column faster than smaller proteins, which will spend time inside the
beads. The proteins gradually become separated and can be recovered by collecting the solution that exits the column (Fig. 4-20). (a) 109 Tools and Techniques: Analyzing Protein Structure (b) Figure 4-20 Size-exclusion chromatography. (a) Small molecules (blue) can enter the spaces inside the porous beads of the stationary phase, while larger
molecules (gold) are excluded. (b) When a mixture of proteins (green) is applied to the top of a size-exclusion column, the large proteins (gold) migrate more quickly than small proteins (blue) through the column and are recovered by collecting the material that flows out the bottom of the column. In this way, a mixture of proteins can be separated
according to size. See Animated Figure. Gel filtration chromatography. A protein’s net charge at a particular pH can be exploited for its purification by ion exchange chromatography. In this technique, the solid phase typically consists of beads derivatized with positively charged diethylaminoethyl (DEAE) groups or negatively charged carboxymethyl
(CM) groups: DEAE CM CH2 CH2 CH2 CH2 NH⪕ CH2 CH3 CH3 COO⪖ Negatively charged proteins will bind tightly to the DEAE groups, while uncharged and positively charged proteins pass through the column. The bound proteins can then be dislodged by passing a high-salt solution through the column so that the dissolved ions can compete with
the protein molecules for binding to DEAE groups (Fig. 4-21). Alternatively, the pH of the solvent can be decreased Figure 4-21 Ion exchange High salt High salt chromatography. When a mixture of proteins (brown) is applied to the top of a positively charged anion exchange column (e.g., a DEAE matrix), negatively charged proteins (red) bind to the
matrix, while uncharged and cationic proteins (orange and purple) flow through the column. The desired protein can be dislodged by applying a high-salt solution (whose anions compete with the protein for binding to the DEAE groups). In cation exchange chromatography, negatively charged proteins bind to an anionic matrix (e.g., one containing
carboxymethyl groups). See Animated Figure. Ion exchange chromatography. so that the bound protein’s anionic groups become protonated, loosening their hold on the DEAE matrix. Similarly, positively charged proteins will bind to CM groups and can subsequently be dislodged by solutions with a higher salt concentration or a higher pH. The
success of ion exchange can be enhanced by knowing something about the protein’s net charge (see Sample Calculation 4-1) or its isoelectric point, pI, the pH at which it carries no net charge. For a molecule with two ionizable groups, the pI lies between the pK values of those two groups: 110 Ch 4 Protein Structure pI 5 1/2 (pK1 1 pK2) [4-1]
Calculating the pI of an amino acid is relatively straightforward (Sample Calculation 4-2). However, a protein may contain many ionizable groups, so although its pI can be estimated from its amino acid composition, its pI is more accurately determined experimentally. SAMPLE CALCULATION 4-2 PROBLEM Estimate the isoelectric point of arginine.
SOLUTION In order for arginine to have no net charge, its a-carboxyl group must be unprotonated (negatively charged), its a-amino group must be unprotonated (neutral), and its side chain must be protonated (positively charged). Because protonation of the a-amino group or deprotonation of the side chain would change the amino acid’s net charge,
the pK values of these groups (9.0 and 12.5) should be used with Equation 4-1: pI 5 1/2 (9.0 1 12.5) 5 10.75 PRACTICE PROBLEMS Figure 4-22 SDS-PAGE. Following electrophoresis, the gel was stained with Coomassie blue dye so that each protein is visible as a blue band. Lanes 1 and 8 contain proteins of known molecular mass that serve as
standards for estimating the masses of other proteins. [Courtesy of Bio-Rad Laboratories, ©2012.] 4. Estimate the pI of alanine. 5. Estimate the pI of glutamate. Other binding behaviors can be adapted for chromatographic separations. For example, a small molecule can be immobilized on the chromatographic matrix, and proteins that can specifically
bind to that molecule will stick to the column while other substances exit the column without binding. This technique, called affinity chromatography, is a particularly powerful separation method because its takes advantage of a protein’s unique ability to interact with another molecule, rather than one of its general features such as size or charge.
High-performance liquid chromatography (HPLC) is the name given to chromatographic separations, often analytical in nature rather than preparative, that are carried out in closed columns under high pressure, with precisely controlled flow rates and automated sample application. Proteins are sometimes analyzed or isolated by electrophoresis, in
which molecules move through a gel-like matrix such as polyacrylamide under the influence of an electric field (Section 3-4). In sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), both the sample and the gel contain the detergent SDS, which binds to proteins to give them a uniform density of negative charges. When the electric
field is applied, the proteins all move toward the positive electrode at a rate depending on their size, with smaller proteins migrating faster than larger ones. After staining, the proteins are visible as bands in the gel (Fig. 4-22). 111 Tools and Techniques: Analyzing Protein Structure Mass spectrometry reveals amino acid sequences A standard
approach to sequencing a protein has several steps: 1. The sample of protein to be sequenced is purified (for example, by chromatographic or other methods) so that it is free of other proteins. 2. If the protein contains more than one kind of polypeptide chain, the chains are separated so that each can be individually sequenced. In some cases, this
requires breaking (reducing) disulfide bonds. 3. Large polypeptides must be broken into smaller pieces (⬍100 residues) that can be individually sequenced. Cleavage can be accomplished chemically, for example, by treating the polypeptide with cyanogen bromide (CNBr), which cleaves the peptide bond on the C-terminal side of Met residues.
Cleavage can also be accomplished with proteases (another term for peptidases) that hydrolyze specific peptide bonds. For example, the protease trypsin cleaves the peptide bond on the C-terminal side of the positively charged residues Arg and Lys: NH⪕ 3 CH2 NH⪕3 Lys (or Arg) CH2 CH2 CH2 CH2 CH2 CH2 O See Guided Exploration. Protein
sequence determination. R O NH CH C NH CH C H2O CH2 O trypsin NH CH C R O⪖ Some commonly used proteases and their preferred cleavage sites are listed in Table 4-4. 4. Each peptide is sequenced. In a classic procedure known as Edman degradation, the N-terminal residue of a peptide is chemically derivatized, cleaved off, and identified; the
process is then repeated over and over so that the peptide’s sequence can be deduced, one residue at a time. Alternatively, each peptide can be sequenced by mass spectrometry. 5. To reconstruct the sequence of the intact polypeptide, a different set of fragments that overlaps the first is generated so that the two sets of sequenced fragments can be
aligned. Set 1 (cleaved with trypsin) Val Leu Lys Ser Phe Gly Arg Tyr Ala Gln Thr Set 2 (cleaved with chymotrypsin) Val Leu Lys Ser Phe Gly Arg Tyr Ala Gln Thr A more efficient approach for analyzing protein structure measures the sizes of peptide fragments by mass spectrometry. In standard mass spectrometry, a solution of the protein (or another
macromolecule) is sprayed from a tiny nozzle at high voltage. This yields droplets of positively charged molecular fragments, from which the solvent quickly evaporates. Each gas-phase ion then passes through an ⪕ ⪕ H3N O CH C TABLE 4-4 Specificities of Some Proteases Protease Residue Preceding Cleaved Peptide Bonda Chymotrypsin Elastase
Thermolysin Trypsin Phe, Trp, Tyr Ala, Gly, Ser, Val Ile, Met, Phe, Trp, Tyr, Val Arg, Lys a Cleavage does not occur if the following residue is Pro. See Animated Figure. Generating overlapping peptides to determine amino acid sequence. 112 Ch 4 Protein Structure MS1 ⪕ Protein ⪕ ⪕ Collision chamber ⪕ ⪕ ⪕ ⪕ ⪕ ⪕ ⪕ ⪕ ⪕ ⪕ ⪕ ⪕ ⪕ Ionization Selected
peptide ⪕ ⪕ MS2 ⪕ ⪕ ⪕ ⪕ ⪕ ⪕ Fragmentation Mass analysis Isolation Val(99.3) (a) 0 200 600 800 1000 1200 1400 Gly(56.9) 1964.8965 Val(99.4) Ala(71.2) Ala(71.2) Ser(86.7) Gly(56.7) Asp(115.1) Ser(86.8) Gly(57.3) Ala(71.4) Ser(86.9) Leu/lle(113.1) 400 Ser(87.2) Ala(71.2) 500 Pro(97.0) 1000 Ser(87.0) Tyr(163.1) Counts 1500 Leu/lle(113.5)
Val(99.3) 2000 1600 1800 Mass (m/z) (b) Figure 4-23 Peptide sequencing by mass spectrometry. (a) A solution of charged peptides is sprayed into the first mass spectrometer (MS1). One peptide ion is selected to enter the collision chamber to be fragmented. The second mass spectrometer (MS2) then measures the mass-to-charge ratio of the ionic
peptide fragments. The peptide sequence is determined by comparing the masses of increasingly larger fragments. (b) An example of peptide sequencing by mass spectrometry. The difference in mass of each successive peak identifies each residue, allowing the amino acid sequence to be read from right to left. [From Keough, T., et al., Proc. Nat.
Acad. Sci. 96, 7131–7136 (1999). Reprinted with permission of PNAS.] electric field. The ions are deflected, with smaller ions deflected more than larger ions, so that they are separated by their mass-to-charge ratios. In this way, the masses of the fragments can be measured and the mass of the intact molecule can be deduced. Two mass
spectrometers in series can be used to determine the sequence of amino acids in a polypeptide. The first instrument sorts the peptide ions so that only one emerges. This species is then allowed to collide with an inert gas, such as helium, which breaks the peptide, usually at a peptide bond. The second mass spectrometer then measures the mass-to-
charge ratios of the peptide fragments (Fig. 4-23). Because successively larger fragments all bear the same charge but differ by one amino acid, and because the mass of each of the 20 amino acids is known, the sequence of amino acids in a set of fragments can be deduced. Although mass spectrometry is not practical for sequencing large
polypeptides, even a partial sequence may be valuable (Box 4-D). Mass spectrometry can also reveal additional details of protein structure, such as covalently modified amino acid residues. BOX 4-D BIOCHEMISTRY NOTE 113 Tools and Techniques: Analyzing Protein Structure Mass Spectrometry Applications Mass spectrometry has been used for
several decades in clinical and forensics laboratories to identify normal metabolic compounds as well as toxins and drugs (both therapeutic and illicit). Instruments that detect traces of explosives at airport security checkpoints also use mass spectrometry, which is rapid, sensitive, and reliable. The analysis of small compounds by mass spectrometry is
relatively straightforward. It is much more challenging, however, to analyze large molecules in complex mixtures for the purpose of diagnosing a disease such as cancer or tracking its effects on human tissues. Fluids such as blood contain so many different proteins, with concentrations ranging over many order of magnitude, that it is difficult to
detect rare proteins against a backdrop of very abundant ones such as serum albumin and immunoglobulins, which together account for about 75% of the plasma proteins. Urinalysis by mass spectrometry holds more promise, as urine normally contains relatively few proteins and none with a mass greater than about 15,000 D. Even so, over 2000
different proteins are detectable in urine. One common approach for identifying proteins in biological samples is to fractionate the mixture by electrophoresis, then extract the separated proteins from the gel, partially digest them with a protease, and subject them to mass spectrometry. The resulting pattern of peaks, a “fingerprint,” can be compared
to a database to identify the protein. Even when tandem mass spectrometers are used, it may not be necessary to completely sequence each polypeptide. A partial sequence of just a few amino acids is often sufficient to identify the parent protein. Of course, the availability of complete genomic sequences makes this approach possible, and a number of
software programs have been developed to translate mass spectral data into query sequences for searching sequence databases. Question: Explain why it is possible that a 5-residue sequence can uniquely identify a protein whose total length is 200 residues. (Hint: compare the total number of possible pentapeptide sequences to the actual number of
pentapeptide fragments in the polypeptide.) Protein structures are determined by X-ray crystallography, electron crystallography, and NMR spectroscopy Most proteins are too small to be directly visualized, even by electron microscopy, but their atomic structures are accessible to high-energy probes in the form of X-rays. X-Ray crystallography is
performed on samples of protein that have been induced to form crystals. A protein preparation must be exceptionally pure in order to crystallize without imperfections. A protein crystal, often no more than 0.5 mm in diameter, usually contains 40–70% water by volume and is therefore more gel-like than solid (Fig. 4-24). When bombarded with a
narrow beam of X-rays, the electrons of the atoms in the crystal scatter the X-rays, which reinforce and cancel each other to produce a diffraction pattern of light and dark spots that can be captured electronically or on a piece of X-ray film (Fig. 4-25). Mathematical analysis of the intensities and positions of the diffracted X-rays yields a three-
dimensional map of electron density in the crystallized molecule. The level of detail of the image depends in part on the quality of the crystal. Slight conformational variations among the crystallized protein molecules often limit resolution to about 2 Å. However, this is usually sufficient to trace the polypeptide backbone and discern the general shapes
of the side chains (Fig. 4-26). When Kendrew determined the X-ray structure of myoglobin, its amino acid sequence was not known. Today, protein crystallographers usually take advantage of amino acid sequence information to simplify the task of elucidating protein structures. Figure 4-24 Crystals of the protein streptavidin. [Courtesy I. Le Trong
and R. E. Stenkamp, University of Washington.] Figure 4-25 An X-ray diffraction pattern. [Courtesy Isolde Le Trong, David Teller, and Ronald Stenkamp, University of Washington.] 114 Ch 4 Protein Structure 5.0 Å 3.0 Å 1.5 Å Figure 4-26 Protein structure at different resolutions. In this example, the green areas represent electron density, and the
protein structure is superimposed in black. As the resolution improves, it becomes easier to trace the pattern of the protein backbone (a portion of an a helix). [Based on a drawing by Wayne Hendrickson.] Figure 4-27 The NMR structure of glutaredoxin. Twenty structures for this 85-residue protein, all compatible with the NMR data, are shown as a-
carbon traces, colored from the N-terminus (blue) to the C-terminus (red). [Structure (pdb 1EGR) determined by P. Sodano, T. H. Xia, J. H. Bushweller, O. Bjornberg, A. Holmgren, M. Billeter, and K. Wuthrich.] Are X-ray structures accurate? One might expect a crystallized molecule to be utterly immobile, quite different from a protein in solution,
which undergoes bending and stretching movements among its many bonds. But, in fact, crystallized proteins retain some of their ability to move. They can sometimes bind small molecules that diffuse into the crystal (which is about half water) and can sometimes mediate chemical reactions. These activities would not be possible if the structure of
the crystallized protein did not closely approximate the structure of the protein in solution. Finally, nuclear magnetic resonance (NMR) methods (see below) for determining the structures of small proteins in solution appear to yield results consistent with X-ray crystallographic data. Proteins that are difficult to crystallize, such as membrane proteins
with extensive hydrophobic regions, can sometimes be analyzed by electron crystallography, a technique in which electron beams rather than X-rays probe the protein’s structure. Protein samples, which do not have to be in a crystalline array, are placed in an electron microscope, and diffraction information is collected from many different angles. In
this way, the three-dimensional structure of the protein can be reconstructed. Because electrons interact more strongly with atoms than X-rays do, the sample is susceptible to radiation damage. This can be minimized by rapidly freezing the sample and collecting data at the temperature of liquid nitrogen (–196°C), a method called cryoelectron
microscopy. The structures of some macromolecular complexes, including the ribosome (Section 22-2), have been visualized using cryoelectron microscopy. Proteins in solution can be analyzed by nuclear magnetic resonance (NMR) spectroscopy, which takes advantage of the ability of atomic nuclei (most commonly, hydrogen) to resonate in an
applied magnetic field according to their interactions with nearby atoms. An NMR spectrum consists of numerous peaks that can be analyzed to reveal the distances between two H atoms that are close together in space or are covalently connected through one or two other atoms. These measurements, along with information about the protein’s
amino acid sequence, are used to construct a threedimensional model of the protein. Due to the inherent imprecision of the measurements, NMR spectroscopy typically yields a set of closely related structures, which may convey a sense of the protein’s natural conformational flexibility (Fig. 4-27). CONCEPT REVIEW • What types of chromatography
would separate proteins of different size? Of different charge? • Summarize the strategy used to determine the sequence of a large protein. • Compare the physical state of the protein to be analyzed by X-ray crystallography, electron crystallography, and NMR spectroscopy. 115 SUMMARY 4-1 Proteins Are Chains of Amino Acids • The 20 amino acid
constituents of proteins are differentiated by the chemical properties of their side chains, which can be roughly classified as hydrophobic, polar, or charged. • Amino acids are linked by peptide bonds to form a polypeptide. • A denatured protein may refold to achieve its native structure. In a cell, chaperones assist protein folding. 4-4 Quaternary
Structure • Proteins with quaternary structure have multiple subunits. 4-2 Secondary Structure: The Conformation of the Peptide Group 4-5 Tools and Techniques: Analyzing Protein Structure • Protein secondary structure includes the a helix and b sheet, in which hydrogen bonds form between backbone carbonyl and amide groups. Irregular
secondary structure has no regularly repeating conformation. • In the laboratory, proteins can be purified by chromatographic techniques that take advantage of the proteins’ size, charge, and ability to bind other molecules. • The sequence of amino acids in a polypeptide can be determined by chemically derivatizing and removing them in Edman
degradation or by measuring the mass-to-charge ratio of peptide fragments in mass spectrometry. • X-Ray crystallography, electron crystallography, and NMR spectroscopy provide information about the three-dimensional structure of proteins at the atomic level. 4-3 Tertiary Structure and Protein Stability • The three-dimensional shape (tertiary structure) of a
protein includes its backbone and all side chains. A protein may contain all a, all b, or a mix of a and b structures. • A globular protein has a hydrophobic core and is stabilized primarily by the hydrophobic effect. Ion pairing, disulfide bonds, and other cross-links may also help stabilize proteins. GLOSSARY TERMS protein polypeptide amino acid R
group a-amino acid Ca chirality disulfide bond condensation reaction peptide bond residue hydrolysis exopeptidase endopeptidase N-terminus C-terminus backbone microenvironment oligopeptide peptide primary structure secondary structure BIOINFORMATICS tertiary structure quaternary structure a helix b sheet parallel b sheet antiparallel b
sheet regular secondary structure irregular secondary structure globular protein fibrous protein domain ion pair zinc finger denaturation renaturation native structure molecular chaperone processing amyloid deposits transmissible spongiform encephalopathy (TSE) PROJECT 3 VISUALIZING THREEDIMENSIONAL PROTEIN STRUCTURES prion
intrinsically unstructured protein subunit homoheterochromatography size-exclusion chromatography ion exchange chromatography pI affinity chromatography HPLC SDS-PAGE protease Edman degradation mass spectrometry X-ray crystallography diffraction pattern electron crystallography cryoelectron microscopy NMR spectroscopy Learn to use
Jmol or PyMOL to view and manipulate structures obtained from the Protein Data Bank. 116 PROBLEMS 4-1 Proteins Are Chains of Amino Acids 1. Amino acids (except for glycine) are chiral; that is, they contain a chiral carbon and thus can form mirror-image isomers, or enantiomers. Biochemists use D and L to distinguish enantiomers instead of the
RS system used by chemists. (a) The structure of L-alanine is shown. Label the chiral carbon and draw the structure of D-alanine. 13. Biochemists sometimes link a recombinant protein to a protein known as green fluorescent protein (GFP), which was first purified from bioluminescent jellyfish. The fluorophore in GFP (shown below) is a derivative of
three consecutive amino acids that undergo cyclization of the polypeptide chain and an oxidation. Identify (a) the three residues and the bonds that result from the (b) cyclization and (c) oxidation reactions. OH COO⪖ H3N C H CH3 H2C (b) The vast majority of proteins in living systems consist of L amino acids. However, D amino acids are found in
some short bacterial peptides that make up cell walls. Why might this be advantageous to the bacteria? 2. How many chiral carbons does threonine have? How many stereoisomers are possible? Draw the stereoisomers of threonine. 3. (a) (b) (c) (d) (e) (f) (g) Which of the 20 standard amino acids are cyclic aromatic sometimes charged at physiological
pH technically not hydrophobic, polar, or charged basic acidic sulfur-containing 4. The side chains of asparagine and glutamine can undergo hydrolysis in aqueous acid. Draw the reactants and products for the deamination reaction. Name the products. 5. Rank the solubility of the following amino acids in water at pH 7: Trp, Arg, Ser, Val, Thr. 6. What
is more soluble, a solution of free histidine or a solution of the histidine tetrapeptide described in Problem 11? 7. At what pH would an amino acid bear both a COOH and an NH2 group? 8. Most amino acids have melting points around 300°C. They are soluble in water but insoluble in nonpolar organic solvents. How do these observations support your
answer to Problem 7? 9. The pK values of the amino and carboxyl groups in free amino acids differ from the pK values of the N- and C-termini of polypeptides. Explain. 10. Histones are basic proteins that bind to DNA. What amino acids are found in abundance in histones and why? What important intermolecular interactions form between histones
and DNA? 11. Estimate the net charge of a His–His–His–His tetrapeptide at pH 6.0. 12. Certain bacteria synthesize cyclic tetrapeptides. (a) Estimate the net charge at pH 7.0 of such a compound that consists of two Pro and two Tyr residues. (b) If the peptide were linear rather than cyclic, what would be its net charge? CH OH L-Alanine NH CH N C N
O O CH2 C 14. Draw the structure of the dipeptide Lys–Glu at pH 7.0. Label the following: (a) peptide bond, (b) N-terminus, (c) C-terminus, (d) an a-amino group and an ε-amino group, (e) an a-carboxylate group and a g-carboxylate group. 15. The artificial non-nutritive sweetener aspartame is a dipeptide with the sequence Asp–Phe. The carboxyl
terminus is methylated. Draw the structure of aspartame at pH 7.0. 16. The zebrafish is a good animal model to use to test the actions of various drugs. The zebrafish produces a novel heptapeptide that binds to opiate receptors in the brain. The sequence of the peptide is Tyr–Gly–Gly–Phe–Met–Gly–Tyr. Draw the structure of the heptapeptide at pH 7.
17. Glutathione is a Cys-containing tripeptide found in red blood cells that functions to remove hydrogen peroxide and organic peroxides that can irreversibly damage hemoglobin and cell membranes. (a) The sequence of glutathione is g-Glu–Cys–Gly. The g-Glu indicates that the first peptide bond forms between the g-carboxyl group of the Glu side
chain and the a-amino group of Cys. Draw the structure of glutathione. (b) Glutathione is sometimes abbreviated as GSH to show the importance of the Cys side chain in reactions. For example, two molecules of GSH react with an organic peroxide as shown below. The glutathione is oxidized to GSSG, and the organic peroxide is reduced to a less
harmful alcohol. Draw the structure of GSSG. 2 GSH 1 ROOOOOH glutathione peroxidase GSSG 1 ROH 1 H2O 18. Why are the terms protein and polypeptide not interchangeable? 19. A fusion protein can be synthesized in which the protein carries a specific tag. These tags are useful in identifying the proteins in various types of experiments. Fusion
proteins with the FLAG epitope bind to specific anti-FLAG antibodies and carry this extra sequence: Asp–Tyr–Lys–Asp–Asp–Asp–Asp–Lys. Draw the structure of the FLAG epitope. 20. The biotech company that sells anti-FLAG epitopes provides the information that fusion proteins carrying the FLAG epitope display the sequence on the surface of the
protein (see Problem 19). Examine the sequence and explain why the company can make this claim. 117 21. You have isolated a tripeptide containing Arg, His, and Pro. How many different sequences are possible for this tripeptide? 22. You have isolated a tetrapeptide with an unknown sequence, and after hydrolyzing its peptide bonds, you recover
Ala, Glu, Lys, and Thr. How many different sequences are possible for this tetrapeptide? 23. Read the following sentences and identify each statement as describing the primary, secondary, tertiary, or quaternary structure of the protein: (a) The shape of myoglobin is roughly spherical. (b) Hemoglobin consists of four polypeptide chains. (c) About one-
third of the amino acid residues in collagen are glycines. (d) The lysozyme molecule contains regions of helical structure. 24. Which level of protein architecture is discussed in Problem 13? 4-2 Secondary Structure: The Conformation of the Peptide Group 25. Draw the structure of the peptide group and draw hydrogen bonds that form between
functional groups of the peptide group and water. 26. The structure of the peptide bond is drawn in Section 4.1 in the trans configuration in which the carbonyl group and the amide hydrogen are on opposite sides of the peptide bond. Draw the structure of the peptide bond in the cis configuration. Which configuration is more likely to be found in
proteins and why? 27. Compare and contrast the structures of the DNA helix and the a helix. 28. The a helix is stabilized by hydrogen bonds that are directed along the length of the helix. The peptide carbonyl group of the nth residue forms a hydrogen bond with the peptide ONH group of the (n 1 4)th residue. Draw the structure of this hydrogen
bond. 29. Proline is known as a helix disrupter; it sometimes appears at the beginning or the end of an a helix but never in the middle. Explain. 30. Interestingly, glycine, like proline (see Problem 29), is not usually found in an a helix, but for the opposite reason. Explain why glycine residues are not usually found in a helices. 31. In 1967, Schiffer and
Edmundson developed a tool called a helical wheel that is still widely used today. A helical wheel is used to visualize an a helix in which the angle of rotation between two consecutive amino acid residues is 100°. Because the a helix consists of 3.6 residues per turn, the pattern repeats after five turns and 18 residues. In the final representation the
view is down the helical axis. 8 1 12 15 5 4 16 11 9 18 2 13 7 6 14 3 10 17 The sequence of a domain of the gp160 protein (found in the envelope of HIV) is shown below, using one-letter codes for the amino acids. Plot this sequence on the helical wheel. What do you notice about the types of amino acid residues on either side of the wheel?
MRVKEKYQHLWRWGWRWG 32. A 24-residue peptide called Pandinin 2 was isolated from scorpion venom and was found to have both antimicrobial and hemolytic properties. The sequence of the first 18 residues of this peptide is shown below. (a) Use the helical wheel shown in Problem 31 to plot the sequence of this peptide. (b) Propose a
hypothesis that explains why the peptide has the ability to lyse bacteria and red blood cells. FWGALAKGALKLIPSLFS 4-3 Tertiary Structure and Protein Stability 33. Examine Figure 4-8 (triose phosphate isomerase). To which of the CATH classes does this protein belong? 34. Choose the amino acid in the following pairs that would be more likely to
appear on the solvent-exposed surface of a protein: (a) Lys or Leu (b) Ser or Ala (c) Phe or Tyr (d) Trp or Gln (e) Asn or Ile 35. In a receptor protein, a Glu residue is altered to an Ala residue in a site-directed mutagenesis experiment. Binding of the ligand to this receptor decreases 100-fold. What does this experiment tell you about the interaction
between the ligand and the receptor? 36. Scientists studying the enzyme PEP carboxylase from corn conducted site-directed mutagenesis studies in which they synthesized three different mutant forms of the enzyme. For each mutant, a lysine was converted to a different amino acid—an Asn, a Glu, or an Arg. Which substitution would you expect
would have the least effect on enzymatic activity? Which substitution would have the greatest effect? Explain. 37. Draw two amino acid side chains that can interact with each other via the following intermolecular interactions: (a) ion pair (b) hydrogen bond (c) van der Waals interaction (London dispersion forces) 38. A type of muscular dystrophy,
called severe childhood autosomal recessive muscular dystrophy (SCARMD), results from a mutation in the gene for a 50-kD muscle protein. The defective protein leads to muscle necrosis. Detailed studies of this protein have revealed that an arginine residue at position 98 has been mutated to a histidine. Why might replacing an arginine with
histidine result in a defective protein? 39. Laboratory techniques for randomly linking together amino acids typically generate an insoluble polypeptide, yet a naturally occurring polypeptide of the same length is usually soluble. Explain. 40. Proteins can be unfolded, or denatured, by agents that alter the balance of weak noncovalent forces that
maintain the native conformation. How would the following agents cause a protein to denature? Be specific about the type of intermolecular forces that would be affected. (a) heat (b) pH 118 (c) amphiphilic detergents (d) reducing agents such as 2-mercaptoethanol (HSCH2CH2OH) 41. In 1957, Christian Anfinsen carried out a denaturation
experiment with ribonuclease. Ribonuclease, a pancreatic enzyme that catalyzes the digestion of RNA, consists of a single chain of 124 amino acids cross-linked by four disulfide bonds. Urea and 2-mercaptoethanol were added to a solution of ribonuclease, which caused it to unfold, or denature. The loss of tertiary structure resulted in a loss of
biological activity. When the denaturing agent (urea) and the reducing agent (mercaptoethanol) were simultaneously removed, the ribonuclease spontaneously folded back up to its native conformation and regained full enzymatic activity in a process called renaturation. What is the significance of this experiment? 42. Insulin consists of two chains, a
shorter A chain and a longer B chain. The two chains are held together with disulfide bonds. In vivo, insulin is processed from proinsulin, a single polypeptide chain. The C chain is removed from proinsulin to form insulin. C chain N P PGELALAQLGGLGGGLEVAGAQP N Q Q A K R E A chain S S G R IVEQCCTSICSLYQLENYCN C R S S A S S K
FVNQHLCGSHLVEALYLVCGERGFFYTP B chain (a) A denaturation/renaturation experiment similar to the one carried out by Anfinsen with ribonuclease (see Problem 41) was carried out using insulin. However, in contrast to Anfinsen’s results, approximately 2–4% of the activity of the native protein was recovered when the urea and 2-
mercaptoethanol were removed by dialysis (this is the level of activity to be expected if the disulfide bridges formed randomly). When the experiment was repeated with proinsulin, 60% of the activity was restored upon renaturation. Explain these observations. (b) The investigators noted that the refolding of proinsulin depended on pH. For example,
when the proinsulin was incubated in pH 7.5 buffer, only 10% of the proinsulin was renatured. But at pH 10.5, 60% of the proinsulin was renatured. Explain these observations. (Hint: The pK value of the cysteine side chain in proinsulin is similar to the pK value of the sulfhydryl group in free cysteine.) 43. In the mid-1980s, scientists noted that if cells
were incubated at 42°C instead of the normal 37°C, the synthesis of a group of proteins dramatically increased. For lack of a better name, the scientists called these heat-shock proteins. It was later determined that the heat-shock proteins were chaperones. Why do you think that the cell would increase its synthesis of chaperones when the
temperature is increased? 44. A protein engineering laboratory studying monoclonal antibody proteins characterized the thermal stability of these proteins by measuring their melting temperature (Tm), which is defined as the temperature at which the proteins are half unfolded. The investigators found that there was a positive correlation between
Tm and the proteins’ OSH molar ratio. In other words, proteins with a higher content of OSH were more thermally stable. Explain the reason for this observation. 45. Spectrin is a protein found in the cytoskeleton of the red blood cell. The cytoskeleton consists of proteins anchored to the cytosolic surface of the membrane and gives the cell the
strength and flexibility to squeeze through capillaries. The spectrin protein chains consist of repeating segments that form a-helical bundles. Recently, a mutant spectrin was isolated in which a glutamine was replaced with proline. What would be the effect of this mutation on the spectrin protein? What are the consequences for the red blood cell? 46.
During evolution, why do insertions, deletions, and substitutions of amino acids occur more often in loops than in elements of regular secondary structure? 4-4 Quaternary Structure 47. The restriction endonucleases EcoRI and EcoRV are dimeric (two-subunit) enzymes (see Section 3-4). Based on how these proteins interact with DNA, do you expect
them to be homo- or heterodimeric? 48. A protein with two identical subunits can often be rotated 180° (halfway) around its axis so as to generate an identical structure; such a protein is said to have rotational symmetry. Why is it not possible for a protein to have mirror symmetry (that is, its halves would be related as if reflected in a mirror)? 49.

Glutathione transferase consists of a dimer of two identical subunits. The dimer is in equilibrium with its constituent monomers. Site-directed mutagenesis studies were carried out on this protein in which two arginine residues were mutated to glutamines and two aspartates were mutated to asparagines. These substitutions caused the equilibrium to shift in favor of the monomeric form of the enzyme. Where were the arginines and aspartates likely to be found on the protein and what is their role in stabilizing the dimeric form of the enzyme? 50. A tetrameric protein dissociates into dimers when the detergent sodium dodecyl sulfate (SDS) is added to a solution of the protein. But the dimers are termed SDS-resistant because they do not further dissociate into monomers in the presence of the detergent. What intermolecular forces might be acting at the dimer– dimer interface? Are the intermolecular forces acting at the monomer– monomer interface different? Explain. 4-5 Tools and Techniques: Analyzing Protein Structure 51. Estimate the pI of a Ser–Ile dipeptide. 52. Estimate the pI of a Gly–Tyr–Val tripeptide. 53. A certain protein has a pI of 4.3. What types of amino acids are likely to be relatively abundant in this protein? 54. What can you conclude about the net charge of any protein at a pH less than 3.5? 55. In your laboratory, you plan to use ion exchange chromatography to separate the peptide shown below (using one-letter codes) from a mixture of different peptides at pH 7.0. Should you choose a matrix containing DEAE or CM groups? Peptide: GLEKSLVRLGDVQPSLGKESRAKKFQRQ 56. The protein allergen from peanuts, a protein called Ara h8, was recently purified and characterized. The investigators initially 119 had difficulty separating Ara h8 from a similar protein in peanuts called Ara h6 because the two proteins were of similar size and had nearly identical pI values. Separation of the two proteins was finally achieved when it was noted that the Ara h6 protein contained 10 cysteine residues involved in disulfide bridges, whereas Ara h8 contained no cysteines. The protein mixture was treated with a reducing agent, dithiothreitol (DTT), and then treated with iodoacetic acid (ICH2COOH, a reagent that adds to, or alkylates, an —SH group and releases free iodine). The mixture was then loaded onto an anion exchange column and the two proteins were successfully separated. (a) Show the structural changes that occur when Cys residues are exposed to DTT followed by iodoacetic acid. (b) Draw a plausible elution profile (protein concentration versus solvent volume) for the separation of the proteins by anion exchange chromatography. (c) Explain why this treatment resulted in successful separation of the two proteins. 57. The sequence of kassinin, a tachykinin dodecapeptide from the African frog Kassina senegalensis, was determined. A single round of Edman degradation identifies Asp as the N-terminus. A second sample of the peptide is treated with chymotrypsin. Two fragments are released with the following amino acid compositions: fragment I (Gly, Leu, Met, Val) and fragment II (Asp2, Gln, Lys, Phe, Pro, Ser, Val). Next, a third sample of peptide is treated with trypsin, which results in two fragments with the following amino acid compositions: fragment III (Asp, Pro, Lys, Val) and fragment IV (Asp, Gln, Gly, Leu, Met, Phe, Ser, Val). Treatment of another sample with elastase yields a single Gly residue and three fragments—fragment V (Leu, Met), fragment VI (Asp, Lys, Pro, Ser, Val), and fragment VII, which was sequenced: Asp– Gln–Phe–Val. A fourth sample was treated with CNBr, but the dodecapeptide was not cleaved. What is the sequence of the dodecapeptide? 58. The ubiquitin protein from the malaria parasite was treated with chymotrypsin, and the resulting fragments were sequenced. A second sample of the polypeptide was treated with trypsin, and the fragments were sequenced. What is the sequence of the polypeptide? Chymotrypsin fragments Trypsin fragments AGKQLEDGRTLSDY IPPDQQRLIF VKTLTGKTITLDVEPSDTIENVKAKIQDKEGI NIQKESTLHLVLRLRGGMQIF LR AK EGI IQDK LIFAGK QLEDGR TLTGK IPPDQQR GGMQIFVK TLSDYNIQK ESTLHLVLR TITLDVEPSDTIENVK 59. Before sequencing, a protein whose two identical polypeptide chains are linked by a disulfide bond must be reduced and alkylated (to chemically block them). Why should reduction and alkylation also be performed for a single polypeptide chain that includes an intramolecular disulfide bond? 60. You must cleave the following peptide into smaller fragments. Which of the proteases listed in Table 4-4 would theoretically yield the most fragments? The fewest? NMTQGRCKPVNTFVHEPLVDVQNVCFKE 61. Seedlings use seed storage proteins as an important nitrogen source during germination. The amino acid sequence of a seed storage protein named BN was determined in the following manner. The BN protein was first treated with 2-mercaptoethanol to reduce any disulfide bridges. This treatment revealed that BN consists of two chains, a light chain and a heavy chain. Next, the two chains were separated and three separate samples of each chain were treated with different proteases. The fragments obtained were individually sequenced by Edman degradation. The sequence is shown below (the first five residues of the light chain are missing due to a blocked amino terminus). Light chain RIPKCRKEFQQAQHLRACQQWLHKQANQSGGGPS Heavy chain PQGPQQRPPLLQQCCNEKHQEEPLCVCPTLKGASKAVRQQIRQQGQQQGQQGQQLQREISRIYQTATHLPRVCNIPRVSICPFQKTMPGP (a) Why was it necessary to carry out a minimum of two different proteolytic cleavages of the protein using different proteases? (b) One of the enzymes used by the investigators was trypsin. Indicate the sequences of the fragments that would result from trypsin digestion. (c) Choose a second protease to cleave both the light and heavy chains into smaller fragments. What protease did you choose, and why? Indicate the sequences of the fragments that would result from digestion by the protease you chose. 62. A peptide sequence linked to the amino terminal of an antibody was sequenced by purifying its mRNA and then using the genetic code to determine the protein sequence. The amino acid sequence of this peptide is shown below: METDTLLLWVLLLWVPGSTG Why would sequencing using the traditional methods (e.g., enzymatic cleavage) be difficult to accomplish with this particular peptide? 63. In prokaryotes, the error rate in protein synthesis may be as high as 5 3 1024 per codon. How many polypeptides containing 500 residues would you expect to contain an amino acid substitution? 64. How many polypeptides containing 2000 residues would you expect to contain an amino acid substitution (see Problem 63)? 65. The mass of each amino acid residue is shown below. Explain why mass spectrometry, which is highly accurate, cannot distinguish Leu and Ile. Residue Ala Arg Asn Asp Cys Gln Glu Gly His Ile Leu Lys Met Mass (D) 71.0 156.1 114.0 115.0 103.0 128.1 129.0 57.0 137.1 113.1 113.1 128.1 131.0 120 Residue Mass (D) Phe Pro Ser Thr Trp Tyr Val 68. For the peptide described in Problem 67, determine the difference in mass between the smallest and the next-smallest fragment. 147.1 97.1 87.0 101.0 186.1 163.1 99.1 69. X-ray crystallographic analysis of a protein crystal sometimes fails to reveal the positions of the first few residues of a polypeptide chain. Explain. 66. In sequencing by mass spectrometry, not every peptide bond may break. (a) If cleavage between two Gly residues does not occur, which amino acid would be identified in place of the two glycines? (b) What amino acid would be identified if a bond between Ser and Val did not break (see Problem 65 for amino acid masses)? 67. The peptide shown here is subjected to mass spectrometry to determined its sequence. If all its peptide bonds (but no others) are broken, what is the mass of the smallest fragment? Assume that the only charged group is the N-terminus. ⩽ H3N H3C O N H O⩾ CH3 O CH3 N H N O C O⩾ C C C C N C C O C C C O C N O O H N 70. X-ray crystallography yields models showing the positions of atoms such as C, N, and O. Hydrogen atoms are not detected in most electron density maps and must be added back in their inferred positions. Add back the H atoms in the polypeptide structure below. N C C O S O SELECTED READINGS Branden, C., and Tooze, J., Introduction to Protein Structure (2nd ed.), Garland Publishing (1999). [A well-illustrated book with chapters introducing amino acids and protein structure, plus chapters on specific proteins categorized by their structure and function.] Cuff, A. L., Sillitoe, I., Lewis, T., Redfern, O. C., Garratt, R., Thornton, J., and Orengo, C. A., The CATH classification revisited—architectures reviewed and new ways to characterize structural divergence in superfamilies, Nuc. Acids Res. 37, D310– D314 (2009). [Summarizes a widely used approach to classifying protein structures.] Goodsell, D. S., Visual methods from atoms to cells, Structure 13, 347–354 (2005). [Discusses the challenges of presenting different features of molecular structures.] Hartl, F. U., Bracher, A., and Hayer-Hartl, M., Molecular chaperones in protein folding and proteostasis, Nature 475, 324–332 (2011). [Includes summaries of protein folding pathways and the fate of misfolded proteins.] Proteopedia, www.proteopedia.org [This interactive resource showcases tutorials on the three-dimensional structures of proteins, nucleic acids, and other molecules.] Richardson, J. S., Richardson, D. C., Tweedy, N. B., Gernert, K. M., Quinn, T. P., Hecht, M. H., Erickson, B. W., Yan, Y., McClain, R. D., Donlan, M. E., and Surles, M. C., Looking at proteins: representations, folding, packing, and design, Biophys. J. 63, 1186–1209 (1992). [A highly readable review showing a variety of ways of visualizing protein folding patterns.] Soto, C., and Estrada, L. D., Protein misfolding and neurodegeneration, Arch. Neurol. 65, 184–189 (2008). [Summarizes the evidence for misfolded proteins as the cause of Alzheimer's and Parkinson's diseases, TSEs, and other disorders.] Street, J. M., and Dear, J. W., The application of massspectrometry-based protein, biomarker discovery to theragnostics, Br. J. Clin. Pharmacol. 69, 367–378 (2010). [Discusses how mass spectrometry can be used to discover and identify proteins associated with diseases.] chapter 5 PROTEIN FUNCTION CAN proteins move? The traditional approach to studying proteins focuses on their architecture, such as a helices and b sheets, suggesting that proteins are like sculpture: beautiful perhaps, but somewhat rocklike. Indeed, a common technique for determining protein structure requires that the protein molecules first form a crystal. The static images of proteins that appear on the pages of this textbook further contribute to the impression that proteins are not particularly dynamic. This chapter highlights several types of proteins, describing how they move—not just by bonds stretching and flexing but also by dramatic movements that are integral to the proteins' physiological functions. The swimming movements of the parasite Giardia, for example, depend on the activity of motor proteins in its flagella. [Alliance Images/Alamy Limited] THIS CHAPTER IN CONTEXT Do You Remember? Part 1 Foundations Part 2 Molecular Structure and Function 5 Protein Function Noncovalent forces, including hydrogen bonds, ionic interactions, and van der Waals forces, act on biological molecules (Section 2-1). A protein's structure may be described at four levels, from primary to quaternary (Section 4-1). Some proteins can adopt more than one stable conformation (Section 4-3). Proteins containing more than one polypeptide chain have quaternary structure (Section 4-4). Part 3 Metabolism Part 4 Genetic Information 121 122 Ch 5 Protein Function Every protein, with its unique three-dimensional structure, can perform some unique function for the organism that produces it. For example, the hormone insulin (a 51-residue protein) docks with its receptor (a much larger membrane protein) to trigger certain intracellular responses. And nearly all enzymes are proteins that interact with specific molecules and mediate their chemical transformation. This chapter begins by examining myoglobin, an intracellular oxygen-binding protein that gives vertebrate muscle a reddish color, and hemoglobin, a major protein of red blood cells, which transports O2 from the lungs to other tissues. These two proteins have been studied for many decades and provide a wealth of information about protein function. Myoglobin and hemoglobin are globular proteins, but many of the most abundant proteins are fibrous proteins that are elongated and may assemble to form extended structures that determine the shape and other physical attributes of cells and entire organisms. These structural proteins include collagen, an extracellular matrix protein, and a variety of proteins that constitute the intracellular scaffolding known as the cytoskeleton. Other than forming fibrous networks, these proteins have little in common, exhibiting a variety of secondary, tertiary, and quaternary structural characteristics related to their distinct physiological functions. Whereas the supportive role of fibrous proteins in cellular architecture may seem obvious, it turns out that many of the dynamic functions of cells are also intricately tied to the cytoskeleton. The movements of cells and the movements of organelles within cells reflect the action of motor proteins that operate along tracks provided by cytoskeletal fibers. Motor proteins provide some additional lessons in protein structure and function due to the mechanism whereby they use chemical energy (the energy of ATP) to do mechanical work, that is, molecular movement. 5-1 Myoglobin and Hemoglobin: Oxygen-Binding Proteins KEY CONCEPTS • O2 binds to the heme group of myoglobin such that binding is half-maximal when the oxygen concentration is equal to the dissociation constant. • The similarities in structure and sequence between myoglobin and hemoglobin indicate a common evolutionary origin. • O2 can bind cooperatively to hemoglobin as the protein shifts from the deoxy to the oxy conformation. • The Bohr effect and BPG modulate hemoglobin function in vivo. Myoglobin is a relatively small protein with a compact shape about 44 3 44 3 25 Å (Fig. 5-1a). Myoglobin lacks b structure entirely, and all but 32 of its 153 amino acids are part of eight a helices, which range in length from 7 to 26 residues and are labeled A through H (Fig. 5-1b). Hemoglobin is a tetrameric protein whose four subunits each resemble myoglobin. The fully functional myoglobin molecule contains a polypeptide chain plus the iron-containing porphyrin derivative known as heme (shown below). The heme is a type of prosthetic group, an organic compound that allows a protein to carry out some function that the polypeptide alone cannot perform—in this case, binding oxygen. ⩾O C O O CH2 CH2 CH3 CH CH2 C C C C C CH C N⩽ N N ⩽N C C C C C CH2 CH2 CH C ⩾O C C Fe CH3 CH3 CH CH C C C CH CH2 CH3 The planar heme is tightly wedged into a hydrophobic pocket between helices E and F of myoglobin. It is oriented so that its two nonpolar vinyl (OCHPCH2) groups are buried and its two polar propionate (OCH2OCH2OCOO2) groups are exposed to the solvent. The central iron atom, with six possible coordination bonds, is liganded by four N atoms of the porphyrin ring system. A fifth ligand is provided by a His residue of myoglobin known as His F8 (the eighth residue of helix F). Molecular oxygen (O2) can bind reversibly to form the sixth coordination bond. (This is what allows certain heme-containing proteins, such as myoglobin and hemoglobin, to function physiologically as oxygen carriers.) Residue His E7 (the seventh residue of helix E) forms a hydrogen bond to the O2 molecule (Fig. 5-2). By itself, heme is not an effective oxygen carrier because the central Fe(II) (or Fe21) atom is easily oxidized to the ferric Fe(III) (or Fe31) state, which cannot bind O2. Oxidation does not readily take place when the heme is part of a protein such as myoglobin or hemoglobin. (a) C F D E Oxygen binding to myoglobin depends on the oxygen concentration B H The muscles of diving mammals are especially rich in myoglobin. At one time, myoglobin was believed to be an oxygen-storage protein—which would be advantageous during a long dive—but it most likely just facilitates oxygen diffusion through muscle cells or binds other small molecules such as nitric oxide (NO). Myoglobin's O2-binding behavior can be quantified. To begin, the reversible binding of O2 to myoglobin (Mb) is described by a simple equilibrium Mb 1 O2 Δ MbO2 with a dissociation constant, K: K5 [Mb][O2] [MbO2] [5-1] where the square brackets indicate molar concentrations. (Note that biochemists tend to describe binding phenomena in terms of dissociation constants, sometimes given as K d, which are the reciprocals of the association constants, K a, used by chemists.) The proportion of the total myoglobin molecules that have bound O2 is called the fractional saturation and is abbreviated Y: Y5 [MbO2] [Mb] 1 [MbO2] G A (b) Figure 5-1 Models of myoglobin structure. (a) Space-filling model. All atoms (except H) are shown (C gray, O red, N blue). The heme group, where oxygen binds, is purple. (b) Ribbon diagram with the eight a helices labeled A–H. [Structure of myoglobin (pdb 1MBO) determined by S. E. V. Phillips.] ? To which class of the CATH system (Section 4-3) does myoglobin belong? [5-2] Since [MbO2] is equal to [Mb][O2]/K (Equation 5-1, rearranged), Y5 [O2] K 1 [O2] [5-3] O2 is a gas, so its concentration is expressed as p O2, the partial pressure of oxygen, in units of torr (where 760 torr 5 1 atm). Thus, Y5 p O2 K 1 p O2 [5-4] In other words, the amount of O2 bound to myoglobin (Y) is a function of the oxygen concentration (pO2) and the affinity of myoglobin for O2 (K). A plot of fractional saturation (Y) versus pO2 yields a hyperbola (Fig. 5-3). As the O2 concentration increases, more and more O2 molecules bind to the heme groups of myoglobin molecules until, at very high O2 concentrations, virtually all the myoglobin molecules have bound O2. Myoglobin is then said to be saturated with oxygen. The oxygen concentration at which myoglobin is half-saturated—that is, the concentration of O2 at which Y is half-maximal—is equivalent to K. For convenience, K is usually called p50, the oxygen pressure at 50% saturation. For human myoglobin, p50 is 2.8 torr (see Sample Calculation 5-1). Figure 5-2 Oxygen binding to the heme group of myoglobin. The central Fe(II) atom of the heme group (purple) is liganded to four porphyrin N atoms and to the N of His F8 below the porphyrin plane. O2 (red) binds reversibly to the sixth coordination site, above the porphyrin plane. Residue His E7 forms a hydrogen bond to O2. 124 Ch 5 Protein Function 1.0 0.8 0.6 Y 0.4 0.2 5 10 15 pO2 (torr) 20 25 30 Figure 5-3 Myoglobin oxygen-binding curve. The relationship between the fractional saturation of myoglobin (Y) and the oxygen concentration (pO2) is hyperbolic. When pO2 5 K 5 2.8 torr, myoglobin is half-saturated (Y 5 0.5). SAMPLE CALCULATION 5-1 PROBLEM Calculate the fractional saturation of myoglobin when pO2 5 1 torr, 10 torr, and 100 torr. SOLUTION Use Equation 5-4 and let K 5 2.8 torr. 1 5 0.26 When pO2 5 1 torr, Y 5 2.8 1 1 10 5 0.78 When pO2 5 10 torr, Y 5 2.8 1 10 100 5 0.97 When pO2 5 100 torr, Y 5 2.8 1 100 PRACTICE PROBLEMS 1. Calculate the fractional saturation of myoglobin when pO2 5 5.6 torr. 2. How would the value of Y in Practice Problem 1 change if K 5 1.4 torr? 3. At what pO2 value would myoglobin be 75% saturated with O2? Myoglobin and hemoglobin are related by evolution Hemoglobin is a heterotetramer containing two a chains and two b chains. Each of these subunits, called a globin, looks a lot like myoglobin. The hemoglobin a chains, the hemoglobin b chains, and myoglobin have remarkably similar tertiary structures (Fig. 5-4). All have a heme group in a hydrophobic pocket, a His F8 that ligands the Fe(II) ion, and a His E7 that forms a hydrogen bond to O2. Somewhat surprisingly, the amino acid sequences of the three globin polypeptides are only 18% identical. Figure 5-5 shows the aligned sequences, with the necessary gaps (for example, the hemoglobin a chain has no D helix). The lack of striking sequence similarities among these proteins highlights an important principle of protein three-dimensional structure: Certain tertiary structures—for example, the backbone folding pattern of a globin polypeptide—can accommodate a variety of amino acid sequences. In fact, many proteins with completely unrelated sequences adopt similar structures. Figure 5-4 Tertiary structures of myoglobin and the a and b chains of hemoglobin. Backbone traces of a globin (blue) and b globin (red) are aligned with myoglobin (green) to show their structural similarity. The heme group of myoglobin is shown in gray. [Structure of hemoglobin (pdb 2HHB) determined by G. Fermi and M. F. Perutz.] ? Which portion of globin structure shows the most variability? The least? Explain. Helix A C B D E Mb G-LSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDEMKASEDLKKHGATVLTALG Hb ␣ V-LSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF-DLSH-----GSAQVKGHGKKVADALT Hb VHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPKVKAHGKKVLGAFS Helix F G H Mb GILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSKHPGDFGADAQGAMNKALELFRKDMASNYKELGFQG Hb ␣ NAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR Hb DGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH Figure 5-5 The amino acid sequences of myoglobin and the hemoglobin a and b chains. The sequence of human myoglobin (Mb) and the human hemoglobin (Hb) chains are written so that their helical segments (bars labeled A through H) are aligned. Residues that are identical in the a and b globins are shaded yellow; residues identical in myoglobin and the a and b globins are shaded blue, and residues that are invariant in all vertebrate myoglobin and hemoglobin chains are shaded purple. The one-letter abbreviations for amino acids are given in Figure 4-2. [After Dickerson, R. E., and Geis, I., Hemoglobin, pp. 68–69, Benjamin/Cummings (1983).] ? Can you identify positions occupied by structurally similar amino acids in all three globins? Clearly, the globins are homologous proteins that have evolved from a common ancestor through genetic mutation (see Section 3-3). The a and b chains of human hemoglobin share a number of residues; some of these are also identical in human myoglobin. A few residues are found in all vertebrate hemoglobin and myoglobin chains. The invariant residues, those that are identical in all the globins, are essential for the structure and/or function of the proteins and cannot be replaced by other residues. Some positions are under less selective pressure to Myoglobin maintain a particular amino acid match and can be conservatively substituted by a similar amino acid (for example, Ile for Leu or Ser for Thr). Still other positions are variable, meaning that they can accommodate a variety of residues, none of which is critical for the protein's structure or function. By looking at the similarities and differences in sequences among evolutionarily related proteins such as the globins, it is possible to deduce considerable information about elements of protein structure that are central to protein function (see Bioinformatics Project 4). Sequence analysis also provides a window on the course of globin evolution, since the number of sequence differences roughly corresponds to the time since the genes diverged. An estimated 1.1 billion years ago, a single globin gene was duplicated, possibly by aberrant genetic recombination, leaving two globin genes that then could evolve independently (Fig. 5-6). Over time, the gene sequences diverged by mutation. One gene became the myoglobin gene. The other coded for a monomeric hemoglobin, which is still found in some primitive vertebrates such as the lamprey (an organism that originated about 425 million years ago). Subsequent duplication of the hemoglobin gene and additional sequence changes yielded the a and b globins, which made possible the evolution of a tetrameric hemoglobin (whose structure is abbreviated a2b2). Additional gene duplications and mutations produced the z chain (from the a chain) and the g and ε chains (from the b chain). In fetal mammals, hemoglobin β δ ε γ ζ α Primordial globin Figure 5-6 Evolution of the globins. Duplication of a primordial globin gene allowed the separate evolution of myoglobin and a monomeric hemoglobin. Additional duplications among the hemoglobin genes gave rise to six different globin chains that combine to form tetrameric hemoglobin variants at various times during development. 126 Ch 5 Protein Function has the composition a2g2, and early human embryos synthesize a z2ε2 hemoglobin. In primates, a recent duplication of the b chain gene has yielded the d chain. An a2d2 hemoglobin occurs as a minor component (about 2%) of adult human hemoglobin. At present, the d chain appears to have no unique biological function, but it may eventually evolve one. Oxygen binds cooperatively to hemoglobin A milliliter of human blood contains about 5 billion red blood cells, each of which is packed with about 300 million hemoglobin molecules. Consequently, blood can carry far more oxygen than a comparable volume of pure water. The oxygen-carrying capacity of the blood can be quickly assessed by measuring the hematocrit (the percentage of the blood volume occupied by red blood cells, which ranges from about 40% (in women) to 45% (in men). Individuals with anemia, too few red blood cells, can sometimes be treated to increase red blood cell production (Box 5-A). The hemoglobin in red blood cells, like myoglobin, binds O2 reversibly, but it does not exhibit the simple behavior of myoglobin. A plot of fractional saturation (Y) versus pO2 for hemoglobin is sigmoidal (S-shaped) rather than hyperbolic (Fig. 5-7). Furthermore, hemoglobin's overall oxygen affinity is lower than that of myoglobin: Hemoglobin is half-saturated at an oxygen pressure of 26 torr (p50 5 26 torr), whereas myoglobin is half-saturated at 2.8 torr. Why is hemoglobin's binding curve sigmoidal? At low O2 concentrations, hemoglobin appears to be reluctant to bind the first O2, but as the pO2 increases, O2 binding increases sharply, until hemoglobin is almost fully saturated. A look at the binding curve in reverse shows that at high O2 concentrations, oxygenated hemoglobin is reluctant to give up its first O2, but as the pO2 decreases, all the O2 molecules are easily given up. This behavior suggests that the binding of the first O2 increases the affinity of the remaining O2-binding sites. Apparently, hemoglobin's four heme groups are not independent but communicate with each other in order to work in a unified fashion. This is known as cooperative binding behavior. In fact, the fourth O2 taken up by hemoglobin binds with about 100 times greater affinity than the first. Hemoglobin's relatively low oxygen affinity and its cooperative binding behavior are the keys to its physiological function (see Fig. 5-7). In the lungs, where the pO2 BOX 5-A BIOCHEMISTRY NOTE Erythropoietin Boosts Red Blood Cell Production Erythropoietin (EPO) is a 165-residue protein hormone that signals the bone marrow to produce more red blood cells (also known as erythrocytes). EPO is synthesized primarily by the kidneys, so individuals with kidney disease are often deficient in EPO and develop anemia. Treating these individuals with recombinant EPO can increase red blood cell production to near-normal levels. The results of EPO treatment have not been lost on athletes, particularly in endurance sports such as cycling; some competitors have found that taking EPO increases their red cell counts, thereby increasing their ability to deliver oxygen to their muscles. The illegal use of EPO to enhance athletic performance is difficult to detect. For one thing, EPO is already present in the body, and an increase in red blood cells by itself is not proof of EPO doping. A clever abuser may even take iron supplements to mask the telltale drop in stored iron that accompanies the increased production of erythrocytes. Successfully detecting the presence of recombinant EPO takes advantage of subtle differences between the natural and recombinant hormones. Although they have the same amino acid sequence, recombinant EPO produced by cultured cells is not derivatized with carbohydrates in the same manner as EPO produced naturally by the kidneys. Questions: Would measurement of mean red blood cell volume or red blood cell number provide the same information as the hematocrit? Why might it be hazardous to take EPO to increase red blood cell levels above normal? Tissues Lungs 1.0 0.8 0.6 Y 0.4 0.2 20 40 60 p O2 (torr) 80 100 Figure 5-7 Oxygen binding to hemoglobin. The relationship between fractional saturation (Y) and oxygen concentration (pO2) is sigmoidal. The pO2 at which hemoglobin is half-saturated (p50) is 26 torr. For comparison, myoglobin's O2-binding curve is indicated by the dashed line. The difference in oxygen affinity between hemoglobin and myoglobin ensures that O2 bound to hemoglobin in the lungs is released to myoglobin in the muscles. This oxygen-delivery system is efficient because the tissue pO2 corresponds to the part of the hemoglobin binding curve where the O2 affinity falls off most sharply. See Animated Figure. Oxygenbinding curve of hemoglobin. is about 100 torr, hemoglobin is about 95% saturated with O2. In the tissues, where the pO2 is only about 20 to 40 torr, hemoglobin's oxygen affinity drops off rapidly (it is only about 55% saturated when the pO2 is 30 torr). Under these conditions, the O2 released from hemoglobin is readily taken up by myoglobin in muscle cells, since myoglobin's affinity for O2 is much higher, even at the lower oxygen pressure. Myoglobin can therefore relay O2 from red blood cells in the capillaries to the muscle cell's mitochondria, where it is consumed in the oxidative reactions that sustain muscle activity. Agents such as carbon monoxide, which interferes with O2 binding to hemoglobin, prevent the efficient delivery of O2 to cells (Box 5-B). BOX 5-B ? Could this O2-delivery system still operate if the p50 for myoglobin was twice as high? BIOCHEMISTRY NOTE Carbon Monoxide Poisoning The affinity of hemoglobin for carbon monoxide is about 250 times higher than its affinity for oxygen. However, the concentration of CO in the atmosphere is only about 0.1 ppm (parts per million by volume), compared to an O2 concentration of about 200,000 ppm. Normally, only about 1% of the hemoglobin molecules in an individual are in the carboxyhemoglobin (Hb CO) form, probably as a result of endogenous production of CO in the body (CO acts as a signaling molecule, although its physiological role is not well understood). Danger arises when the fraction of carboxyhemoglobin rises, which can occur when individuals are exposed to high levels of environmental CO. For example, the incomplete combustion of fuels, as occurs in gas-burning appliances and vehicle engines, releases CO. The concentration of CO can rise to about 10 ppm in these situations and to as high as 100 ppm in highly polluted urban areas. The concentration of carboxyhemoglobin may reach 15% in some heavy smokers, although the symptoms of CO poisoning are usually not apparent. CO toxicity, which occurs when the concentration of carboxyhemoglobin rises above about 25%, causes neurological impairment, usually dizziness and confusion. High doses of CO, which cause carboxyhemoglobin levels to rise above 50%, can trigger coma and death. When CO is bound to some of the heme groups of hemoglobin, O2 is not able to bind to those sites because its low affinity means that it cannot displace the bound CO. In addition, the carboxyhemoglobin molecule remains in a high-affinity conformation, so that even if O2 does bind to some of the hemoglobin heme groups in the lungs, O2 release to the tissues is impaired. The effects of mild CO poisoning are largely reversible through the administration of O2. But because the CO remains bound to hemoglobin with a half-life of several hours, recovery is slow. Question: Sketch the oxygen-binding curves of hemoglobin and carboxyhemoglobin [Hb (CO)2]. 127 128 Ch 5 Protein Function CAN proteins move? A conformational shift explains hemoglobin's cooperative behavior The four heme groups of hemoglobin must be able to sense one another's oxygen-binding status so that they can bind or release their O2 in concert. But the four heme groups are 25 to 37 Å apart, too far for them to communicate via an electronic signal. Therefore, the signal must be mechanical. In a mechanism worked out by Max Perutz, the four globin subunits undergo conformational changes when they bind O2. This is one example of protein structural changes that provide an answer to the question posed at the start of the chapter: Proteins can move. In deoxyhemoglobin (hemoglobin without any bound O2), the heme Fe ion has five ligands, so the porphyrin ring is somewhat dome-shaped and the Fe lies about 0.6 Å out of the plane of the porphyrin ring. As a result, the heme group is bowed slightly toward His F8 (Fig. 5-8). When O2 binds to produce oxyhemoglobin, the Fe—now with six ligands—moves into the center of the porphyrin plane. This movement of the Fe ion pulls His F8 farther toward the heme group, and this in turn drags the entire F helix so that it moves up to 1 Å. The F helix cannot move in this manner unless the entire protein alters its conformation, culminating in the rotation of one ab unit relative to the other. Consequently, hemoglobin has two quaternary structures, corresponding to the oxy and deoxy states. α β β α β α Deoxy α β Oxy The shift in conformation between the oxy and deoxy states primarily involves rotation of one ab unit relative to the other. Oxygen binding decreases the size of the central cavity between the four subunits and alters some of the contacts between subunits. The two conformational states of hemoglobin are Figure 5-8 Conformational changes in hemoglobin upon O2 binding. In deoxyhemoglobin (blue), the porphyrin ring is slightly bowed down toward His F8 (shown in ball-and-stick form). The remainder of the F helix is represented by its alpha carbon atoms. In oxyhemoglobin (purple), the heme group becomes planar, pulling His F8 and its attached F helix upward. The bound O2 is shown in red. See Animated Figure. Movements of heme and F helix in hemoglobin. Thr Thr Thr His His Pro Thr Pro (a) Figure 5-9 Some of the subunit interactions in hemoglobin. The interactions between the ab units of hemoglobin include contacts between side chains. The relevant residues are shown in space-filling form. (a) In deoxyhemoglobin, a His residue on the b chain (blue, left) fits between a Pro and a Thr residue on the a chain (green, right). (b) Upon oxygenation, the His (b) residue moves between two Thr residues. An intermediate conformation (between the deoxy and oxy conformations) is disallowed in part because the highlighted side chains would experience strain. [Structure of human deoxyhemoglobin (pdb 2HHB) determined by G. Fermi and M. F. Perutz; structure of human oxyhemoglobin (pdb 1HHO) determined by B. Shaanan.] formally known as T (for "tense") and R (for "relaxed"). The T state corresponds to deoxyhemoglobin, and the R state corresponds to oxyhemoglobin. Deoxyhemoglobin is reluctant to bind the first O2 molecule because the protein is in the deoxy (T) conformation, which is unfavorable for O2 binding (the Fe atom lies out of the heme plane). However, once O2 has bound, probably to the a chain in each ab pair, the entire tetramer switches to the oxy (R) conformation as the Fe atom and the F helix move. An intermediate conformation is not possible, because the contacts between the ab units do not allow it (Fig. 5-9). Molecular dynamics studies suggest that hemoglobin does not instantaneously snap from one conformation to the other but instead undergoes fluctuations in tertiary structure that precede the shift in quaternary structure. Subsequent O2 molecules bind with higher affinity because the protein is already in the oxy (R) conformation, which is favorable for O2 binding. Similarly, oxyhemoglobin tends to retain its bound O2 molecules until the oxygen pressure drops significantly. Then some O2 is released, triggering the change to the deoxy (T) conformation. This decreases the affinity of the remaining bound O2 molecules, making it easier for hemoglobin to unload its bound oxygen. Because measurements of O2 binding reflect the average behavior of many individual hemoglobin molecules, the result is a smooth curve (as shown in Fig. 5-7). Hemoglobin and many other proteins with multiple binding sites are known as allosteric proteins (from the Greek allos, meaning "other," and stereos, meaning "space"). In these proteins, the binding of a small molecule (called a ligand) to one site alters the ligand-binding affinity of the other sites. In principle, the ligands need not be identical, and their binding may either increase or decrease the binding activity of the other sites. In hemoglobin, the ligands are all oxygen molecules, and O2 binding to one subunit of the protein increases the O2 affinity of the other subunits. H1 ions and bisphosphoglycerate regulate oxygen binding to hemoglobin in vivo Decades of study have revealed the detailed chemistry behind hemoglobin's activity (and have also revealed how molecular defects can lead to disease; Box 5-C). The conformational change that transforms deoxyhemoglobin to oxyhemoglobin alters the microenvironments of several ionizable groups in the protein, including the two N-terminal amino groups of the a subunits and the two His residues near the 129 BOX 5-C CLINICAL CONNECTION Hemoglobin Mutations Mutations in the DNA sequence for the genes that encode the a and b chains of hemoglobin produce hemoglobin proteins with altered amino acid sequences. In some cases, the mutation is benign and the hemoglobin molecules function more or less normally. But in other cases, the mutation results in serious physical complications for the individual, as the ability of the mutant hemoglobin to deliver oxygen to cells is compromised. Mutant hemoglobins are often unstable, which may also result in destruction of the red blood cell (anemia). Hundreds of hemoglobin variants are known, and about 5% of the world's population carries an inherited disorder of hemoglobin. One of the betterknown hemoglobin variants is sickle cell hemoglobin (known as hemoglobin S or Hb S). Individuals with two copies of the defective gene develop sickle cell anemia, a debilitating disease that predominantly affects populations of African descent. The discovery of the molecular defect that causes sickle cell anemia was a groundbreaking event in biochemistry. The disease was first described in 1910, but for many years there was no direct evidence that sickle cell anemia—or any genetic disease— was the result of an alteration in the molecular structure of a protein. Then in 1949, Linus Pauling (who was already on his way to discovering the a helix) showed that hemoglobin from patients with sickle cell anemia had a different electrical charge than hemoglobin from healthy individuals. Eight years later, in 1957, Vernon Ingram identified a single amino acid difference: Glu at position A3 in the b chain is replaced by Val in sickle cell hemoglobin. This was the first evidence that an alteration in a gene caused an alteration in the amino acid sequence of the corresponding polypeptide. The mutation is described in Section 3-2. In normal hemoglobin, the switch from the oxy to the deoxy conformation exposes a hydrophobic patch on the protein surface between the E and F helices. The hydrophobic Val residues on hemoglobin S are optimally positioned to bind to this patch. This intermolecular association leads to the rapid aggregation of hemoglobin S molecules to form long rigid fibers. These fibers physically distort the red blood cell, producing the familiar sickle shape. Because hemoglobin S aggregation occurs In this model of polymerized only among deoxyhemohemoglobin S molecules, the globin S molecules, sickling heme groups are red and the tends to occur when the red mutant Val residues are blue. blood cells pass through [From W. Royer and D. Harrington, J. Mol. Biol. 272, 398–407 (1992).] oxygen-poor capillaries. The misshapen cells can obstruct blood flow and rupture, leading to the intense pain, organ damage, and loss of red blood cells that characterize the disease. Normal red blood cells (top) and a sickled cell (bottom). [From Andrew Skred/Science Photo Library/Photo Researchers and Jacki Lewin, Royal Free Hospital/Science Photo Library/Photo Researchers.] The high frequency of the gene for sickle cell anemia (that is, the mutated b globin gene) was at first puzzling: Genes that lead to disabling diseases tend to be rare because individuals with two copies of the gene usually die before they can pass the gene to their offspring. However, carriers of the sickle cell variant appear to have a selective advantage. They are protected against malaria, a parasitic disease that afflicts about 225 million people and kills about 1 million each year, mostly children. In fact, the sickle cell hemoglobin variant is common in regions of the world where malaria is endemic. In heterozygotes (individuals with one normal and one defective b globin gene), only about 2% of red blood cells undergo sickling. Sickling does not directly lead to elimination of Plasmodium falciparum, the intracellular protozoan that causes malaria. However, the presence of hemoglobin S triggers production of heme oxygenase, an engyme whose reaction products protect against the damage caused by the parasite. The thalassemias result from genetic defects that reduce the rate of synthesis of the a or b globin chains. These disorders are prevalent in the Mediterranean area (the name comes from the Greek word thalassa, meaning "sea") and in South Asia. Depending on the nature of the mutation, individuals with thalassemia may experience mild to severe anemia and their red blood cells may be smaller than normal. But like heterozygotes for hemoglobin S, individuals with thalassemia are apparently more resistant to malaria. The accompanying table lists some hemoglobin residues that are critical for normal function; their mutation produces clinical symptoms. Note that defective hemoglobins are usually named for the place where they were first observed or characterized. Chain Position a 44 a Amino Acid Role Significance Pro Participates in the formation of the a1b2 interface in the deoxy form but not the oxy form. Stabilizes the deoxy form. 141 (C-terminus) Arg Its COO2 forms an ion pair with Lys 127 and its side chain forms an ion pair with Asp 126 in the deoxy form. Stabilizes the deoxy form. b 82 Lys Forms an ion pair to BPG in the central cavity. Stabilizes the deoxy form. b 146 His The side-chain imidazole ring forms an ion pair with Asp 94. It also forms an ion pair with BPG in the central cavity. Stabilizes the deoxy form. Questions: 1. In the mutant hemoglobin Hb Ohio (b142Ala S Asp), the substitution of Asp for Ala results in the displacement of the G helix relative to the H helix in the b chain. This decreases the stability of the b146His–b94Asp ion pair. Draw an oxygen-binding curve that compares the relative p50 values of normal hemoglobin (Hb A) and Hb Ohio. What is the effect of the decreased stability of the His–Asp ion pair on Hb Ohio? 2. Investigators of mutant human hemoglobins often subject the proteins to cellulose acetate electrophoresis at pH 8.5, a pH at which most hemoglobins are negatively charged. The proteins are applied at the negative pole and migrate toward the positive pole when the current is turned on. The more negatively charged the hemoglobin, the faster it migrates to the positive pole. Results for normal (Hb A) and sickle cell hemoglobin (Hb S) are shown. Draw a diagram that shows the results of electrophoresis of Hb A and Hb Ohio (b 142Ala S Asp). Hb S ⩾ ⩽ Hb A 3. Hb Milledgeville (a44Pro S Leu) results in a mutated hemoglobin with altered oxygen affinity. Explain how the oxygen affinity is altered. 4. Hb Providence (b44Lys S Asn) results from a single point mutation of the DNA. (a) What is the change in the DNA that occurred to produce the mutant hemoglobin? (b) Compare the oxygen affinities of Hb Providence and Hb A. (c) There are actually two forms of Hb Providence in affected individuals. Hb Providence in which the b44Lys has been replaced with Asn is referred to as Hb Providence Asn. This mutant hemoglobin can undergo deamidation to produce Hb Providence Asp. Draw the reaction that converts Hb Providence Asn to Hb Providence Asp. (d) Draw a diagram that shows the cellulose acetate electrophoresis results for Hb A, Hb Providence Asn, and Hb Providence Asp. (e) Compare the oxygen affinities of Hb Providence Asp and Hb Providence Asn. 5. Hb Syracuse (b146His S Pro) is a mutant hemoglobin with altered oxygen affinity. (a) Draw a diagram showing the cellulose acetate electrophoresis results for Hb Syracuse and Hb A. (b) Evaluate the ability of Hb Syracuse to respond to normal allosteric effectors of hemoglobin. How is the oxygen affinity affected as a result? C-terminus of the b subunits. As a result, these groups become more acidic and release H1 when O2 binds to the protein: Hb ? H1 1 O2 Δ Hb ? O2 1 H1 Therefore, increasing the pH of a solution of hemoglobin (decreasing [H1]) favors O2 binding by "pushing" the reaction to the right, as written above. Decreasing the pH (increasing [H1]) favors O2 dissociation by "pushing" the reaction to the left. The reduction of hemoglobin's oxygen-binding affinity when the pH decreases is known as the Bohr effect. The Bohr effect plays an important role in O2 transport in vivo. Tissues release CO2 as they consume O2 in respiration. The dissolved CO2 enters red blood cells, where it is rapidly converted to bicarbonate (HCO2 3) by the action of the enzyme carbonic anhydrase (see Box 2-D): 1 CO2 1 H2O Δ HCO2 3 1 H 131 132 Ch 5 Protein Function LUNGS BLOOD O2 Hb O2 TISSUES Hb Hb Hb H⩽ Hb H⩽ H⩽ CO2 O2 Hb O2 H⩽ HCO3⩾ HCO3⩾ CO2 Figure 5-10 Oxygen transport and the Bohr effect. Hemoglobin picks up O2 in the lungs. In the tissues, H1 derived from the metabolic production of CO2 decreases hemoglobin's affinity for O2, thereby promoting O2 release to the tissues. Back in the lungs, hemoglobin binds more O2, releasing the protons, which recombine with bicarbonate to re-form CO2. See Animated Figure. The Bohr effect. ? Write the net equation for the process shown in the diagram. The H1 released in this reaction induces hemoglobin to unload its O2 (Fig. 5-10). In the lungs, the high concentration of oxygen promotes O2 binding to hemoglobin. This causes the release of protons that can then combine with bicarbonate to re-form CO2, which is breathed out. Red blood cells use one additional mechanism to fine-tune hemoglobin function. These cells contain a three-carbon compound, 2,3-bisphosphoglycerate (BPG): ⩾ O 1.0 0.8 Without BPG C O H C OPO23⩾ H C OPO23⩾ H 2,3-Bisphosphoglycerate (BPG) 0.6 Y With BPG 0.4 0.2 20 40 pO2 (torr) 60 Figure 5-11 Effect of BPG on hemoglobin. BPG binds to deoxyhemoglobin but not to oxyhemoglobin. It therefore reduces hemoglobin's O2 affinity by stabilizing the deoxy conformation. See Animated Figure. Effect of BPG and CO2 on hemoglobin. BPG binds in the central cavity of hemoglobin, but only in the T (deoxy) state. The five negative charges in BPG interact with positively charged groups in deoxyhemoglobin; in oxyhemoglobin, these cationic groups have moved away and the central cavity is too narrow to accommodate BPG. Thus, the presence of BPG stabilizes the deoxy conformation of hemoglobin. Without BPG, hemoglobin would bind O2 too tightly to release it to cells. In fact, hemoglobin stripped of its BPG in vitro exhibits very strong O 2 affinity, even at low pO 2 (Fig. 5-11). The fetus takes advantage of this chemistry to obtain O2 from its mother's hemoglobin. Fetal hemoglobin has the subunit composition a2g2. In the g chains, position H21 is not His (as it is in the mother's b chain) but Ser. His H21 bears one of the positive charges important for binding BPG in adult hemoglobin. The absence of this interaction in fetal hemoglobin reduces BPG binding. Consequently, hemoglobin in fetal red blood cells has a higher O2 affinity than adult hemoglobin, which helps transfer O2 from the maternal circulation across the placenta to the fetus. CONCEPT REVIEW • Why does myoglobin require a prosthetic group? • What is the relationship between myoglobin's fractional saturation and the oxygen concentration? • How can the sequences of homologous proteins provide information about residues that are essential or nonessential for a protein's function? • How do the different O2 affinities of myoglobin and hemoglobin relate to their physiological roles? • Explain the structural basis for hemoglobin's cooperative O2-binding behavior. • How do the Bohr effect and BPG regulate O2 transport in vivo? 133 Structural Proteins 5-2 Structural Proteins A typical eukaryotic cell contains three types of cytoskeletal proteins that form fibers extending throughout the cell (Fig. 5-12): These are microfilaments (actin filaments, with a diameter of about 70 Å), intermediate filaments (with a diameter of about 100 Å), and microtubules (with a diameter of about 240 Å). In large multicellular organisms, fibers of the protein collagen provide structural support extracellularly. Bacterial cells also contain proteins that form structures similar to microfilaments and microtubules. In the following discussion, note how the structure of each protein influences the overall structure and flexibility of the fiber as well as the fiber's ability to disassemble and reassemble. Microfilaments are made of actin KEY CONCEPTS • Globular actin subunits associate in a double chain to form a microfilament. • The growth and regression of actin filaments can change a cell's shape. • Microtubules are hollow tubes built from tubulin dimers. • Intermediate filaments are longlasting fibrous proteins consisting of coiled a helices. • Three left-handed Gly-rich helical polypeptides form the collagen triple helix. A major portion of the eukaryotic cytoskeleton consists of microfilaments, or polymers of actin. In many cells, a network of actin filaments supports the plasma membrane and therefore determines cell shape (see Fig. 2-7 and Fig. 5-12). Certain proteins cross-link individual actin polymers to help form bundles of microfilaments, thereby increasing their strength. Monomeric actin is a globular protein with about 375 amino acids (Fig. 5-13). On its surface is a cleft in which adenosine triphosphate (ATP) binds. The adenosine group slips into a pocket on the protein, and the ribose hydroxyl groups and the phosphate groups form hydrogen bonds with the protein. Actin filaments Intermediate filaments Microtubules Figure 5-12 Distribution of cytoskeletal fibers in a single and microtubules. [Courtesy J. Victor Small, Austrian Academy of cell. To make these micrographs, each type of fiber was labeled Sciences, Vienna, Austria.] with a fluorescent probe that binds specifically to one type of cytoskeletal protein. Note how the distribution of actin filaments differs somewhat from that of intermediate filaments ? What type of fiber best defines the cell's nucleus? Polymerized actin is sometimes referred to as F-actin (for filamentous actin, to distinguish it from G-actin, the globular monomeric form). The actin polymer is actually a double chain of subunits in which each subunit contacts four neighboring subunits (Fig. 5-14). Each actin subunit has the same orientation (for example, all the nucleotide-binding sites point up in Fig. 5-14), so the assembled fiber has a distinct polarity. The end with the ATP site is known as the (2) end, and the opposite end is the (1) end. Initially, polymerization of actin monomers is slow because actin dimers and trimers are unstable. But once a longer polymer has formed, subunits add to both ends. Addition is usually much more rapid at the (1) end (hence its name) than at the (2) end (Fig. 5-15). Actin polymerization is driven by the hydrolysis of ATP (splitting ATP by the addition of water) to produce ADP 1 inorganic phosphate (Pi): 134 Ch 5 Protein Function NH2 Figure 5-13 Actin monomer. This protein assumes a globular shape with a cleft where ATP (green) binds. [Structure N N of rabbit actin (pdb 1J6Z) determined by L. R. Otterbein, P. Graceffa, and R. Dominguez.] O ⩾O P O O⩾ P O O⩾ P N N O O O CH2 O⩾ H (−) end ⩽ H2O O H H OH OH H Adenosine triphosphate (ATP) NH2 N N O HO P O⩾ O P O CH2 O⩾ H Figure 5-14 Model of an actin filament. The structure of F-actin was determined from X-ray diffraction data and computer model-building. Fourteen actin subunits are shown (all are different colors except the central actin subunit, whose two halves are blue and gray). [Courtesy Ken Holmes, Max Planck Institute for Medical Research.] O ⩽ O ⩾O P OH O⩾ H H OH OH Adenosine diphosphate (ADP) (+) end N N O H Inorganic phosphate (Pi) This reaction is catalyzed by F-actin but not by G-actin. Consequently, most of the actin subunits in a microfilament contain bound ADP. Only the most recently added subunits still contain ATP. Because ATP-actin and ADP-actin assume slightly different conformations, proteins that interact with microfilaments may be able to distinguish rapidly polymerizing (ATP-rich) microfilaments from longer-established (ADP-rich) microfilaments. Actin filaments continuously extend and retract Microfilaments are dynamic structures. Polymerization of actin monomers is a reversible process, so the polymer undergoes constant shrinking and growing as subunits add to and dissociate from one or both ends of the microfilament (see Fig. 5-15). When the net rate of addition of subunits to one end of a microfilament matches the net rate of removal of subunits at the other end, the polymer is said to be treadmilling (Fig. 5-16). (−) end (+) end 135 Structural Proteins ATP-binding cleft Figure 5-15 Actin filament assembly. A microfilament grows as subunits add to its ends. Subunits usually add more rapidly to the (1) end, which therefore grows faster than the (2) end. Actual microfilaments are much longer than depicted here. (a) Calculations suggest that under cellular conditions, the equilibrium between monomeric actin and polymeric actin favors the polymer. However, the growth of microfilaments in vivo is limited by capping proteins that bind to and block further polymerization at the (1) or (2) ends. A process that removes a microfilament cap will target growth to the uncapped end. New microfilament growth can also occur as branches form along existing microfilaments. A supply of actin monomers to support microfilament growth in one area must come at the expense of microfilament disassembly elsewhere. In a cell, certain proteins sever microfilaments by binding to a polymerized actin subunit and inducing a small structural change that weakens actin–actin interactions and thereby increases the likelihood that the microfilament will break at that point. Actin subunits can then dissociate from the newly exposed ends unless they are subsequently capped. Capping, branching, and severing proteins, along with other proteins whose activity is sensitive to extracellular signals, regulate the assembly and disassembly of actin filaments. A cell containing a network of actin filaments can therefore change its shape as the filaments lengthen in one area and regress in another. Certain cells use this system to move. When a cell crawls along a surface, actin polymerization extends its "leading" edge, while depolymerization helps retract its "trailing" edge (Fig. 5-17a). The high density of growing filament ends at the leading edge of the cell (Fig. 5-17b) illustrates how the rapid formation and outward extension of actin filaments can modulate cell shape and drive cell locomotion. Not only do actin filaments provide structural support and generate cell movement by assembly and disassembly, they also participate in generating tensile force. This system is well Figure 5-16 Actin filament treadmilling. Net assembly at one end balances net dissociation at the other end. (b) Figure 5-17 Actin filament dynamics in cell crawling. (a) Scanning electron micrograph of crawling cells. The leading edges of the cells (lower left) are ruffled where they have become detached from the surface and are in the process of extending. The trailing edges or tails of the cells, still attached to the surface (upper right), are gradually pulled toward the leading edge. The rate of actin polymerization is greatest at the leading edge. [Courtesy Guenter Albrecht-Buehler.] (b) Organization of actin filaments in a fish epithelial cell. At the leading edge of the cell (top), the filaments form a dense and highly branched network. Deeper within the cell (bottom), the filaments are sparser. [Courtesy Tatyana Svitkina, Northwestern University Medical School.] 136 Ch 5 Protein Function developed in muscle cells, where actin filaments are an essential part of the contractile apparatus (see Section 5-3). Tubulin forms hollow microtubules Like actin filaments, microtubules are cytoskeletal fibers built from small globular protein subunits. Consequently, they share with actin filaments the ability to assemble and disassemble on a time scale that allows the cell to rapidly change shape in response to external or internal stimuli. Compared to a microtubule, however, an actin filament is a thin and flexible rod. A microtubule is about three times thicker and much more rigid because it is constructed as a hollow tube. Consider the following analogy: A metal rod with the dimensions of a pencil is easily bent. The same quantity of metal, fashioned into a hollow tube with a larger diameter but the same length, is much more resistant to bending. Figure 5-18 Structure of b-tubulin. The strands of the two b sheets are shown in blue, and the 12 a helices that surround them are green. [Structure of pig tubulin (pdb 1TUB) determined by E. Nogales and K. H. Downing.] Figure 5-19 The tubulin dimer. The guanine nucleotide (gold) in the a-tubulin subunit (bottom) is inaccessible in the dimer, whereas the nucleotide in the b-tubulin subunit (top) is more exposed to the solvent. Bicycle frames, plant stems, and bones are built on this same principle. Cells use hollow microtubules to reinforce other elements of the cytoskeleton (see Fig. 5-12), to construct cilia and flagella, and to align and separate pairs of chromosomes during mitosis. The basic structural unit of a microtubule is the protein tubulin. Two monomers, known as a-tubulin and b-tubulin, form a dimer, and a microtubule grows by the addition of tubulin dimers. Each tubulin monomer contains about 450 amino acids, 40% of them identical in a- and b-tubulin. The core of tubulin consists of a four-stranded and a six-stranded b sheet surrounded by 12 a helices (Fig. 5-18). Each tubulin subunit includes a nucleotide-binding site. Unlike actin, tubulin binds a guanine nucleotide, either guanosine triphosphate (GTP) or its hydrolysis product, guanosine diphosphate (GDP). When the dimer forms, the a-tubulin GTP-binding site becomes buried in the interface between the monomers. The nucleotide-binding site in b-tubulin remains exposed to the solvent (Fig. 5-19). After the tubulin dimer is incorporated into a microtubule and another dimer binds on top of it, the b-tubulin nucleotide-binding site is also sequestered from solvent. The GTP is then hydrolyzed, but the resulting GDP remains bound to b-tubulin because it cannot diffuse away (the GTP in the a-tubulin subunit remains where it is and is not hydrolyzed). Assembly of a microtubule begins with the end-to-end association of tubulin dimers to form a short linear protofilament. Protofilaments then align side-to-side in a curved sheet, which wraps around on itself to form a hollow tube of 13 protofilaments (Fig. 5-20). The microtubule extends as tubulin dimers add to both ends. Like a microfilament, the microtubule is polar and one end grows more rapidly. The (⩽) end, terminating in -tubulin, grows about twice as fast as the (⩾) or ␣-tubulin end because tubulin dimers bind preferentially to the (⩽) end. Disassembly of a microtubule also takes place at both ends but occurs more rapidly at the (1) end. Under conditions that favor depolymerization, the ends of the microtubule appear to fray (Fig. 5-21). This suggests that tubulin dimers do not simply dissociate individually from the microtubule ends but that the interactions between protofilaments weaken before the tubulin dimers come loose. Under certain conditions, microtubule treadmilling can occur when tubulin subunits add to the (1) end as fast as they leave the (2) end. In vivo, the (2) ends are often anchored to some sort of organizing center in the cell. This means that most (a) 137 Structural Proteins (+) end β-Tubulin (b) α-Tubulin Protofilament (−) end Microtubule Figure 5-20 Assembly of a microtubule. (a) ab Dimers of view of a microtubule. [Courtesy Kenneth Downing, Lawrence Berkeley tubulin initially form a linear protofilament. Protofilaments associate side by side, ultimately forming a tube. Tubulin dimers can add to either end of the microtubule, but growth is about twice as fast at the (1) end. (b) Cryoelectron microscopy National Laboratory.] ? Compare a microtubule and an actin filament (Fig. 5-14) in terms of strength and speed of assembly. microtubule growth and regression occur at the (1) end. Microtubule dynamics are also regulated by proteins that cross-link microtubules and promote or prevent depolymerization. Some drugs affect microtubules Compounds that interfere with microtubule dynamics can have drastic physiological effects. One reason is that during mitosis, chromosomes separate along a spindle made of microtubules (Fig. 5-22). The drug colchicine, a product of the meadow saffron plant, causes microtubules to depolymerize, thereby blocking cell division. OCH3 H3CO H3CO O NH H3CO CH3 O Colchicine Colchicine binds at the interface between a- and b-tubulin in a dimer, facing the inside of the microtubule cylinder. The bound drug may induce a slight conformational change that weakens the lateral contacts between protofilaments. If enough colchicine is present, microtubules shorten and eventually disappear. Colchicine was first used over 2000 years ago to treat gout (inflammation stemming from the precipitation of uric acid in the joints) because it inhibits the action of the white blood cells that mediate inflammation. Figure 5-21 Electron micrograph of a depolymerizing microtubule. The ends of protofilaments apparently curve away from the microtubule and separate before tubulin dimers dissociate. [Courtesy Ronald Milligan, The Scripps Research Institute.] 138 Ch 5 Protein Function Taxol binds to b-tubulin subunits in a microtubule, but not to free tubulin, so it stabilizes the microtubule, preventing its depolymerization. O H3C O H3C O C NH O OH CH3 O C CH3 CH3 O O OH OH O O C CH3 O C O Taxol Figure 5-22 Microtubules in a dividing cell. During mitosis, microtubules (green fluorescence) link replicated chromosomes (blue fluorescence) to two points at opposite sides of the cell. Disassembly of the microtubules, along with the action of motor proteins, draws the chromosomes apart before the cell splits in half. [Courtesy Alexey Khodjakov and Conly L. Rieder, Wadsworth Center, Albany.] Figure 5-23 Scanning electron micrograph of sectioned human skin. The layers of dead epidermal cells at the top consist mostly of keratin. [Science Photo Library/Photo Researchers.] The taxol–tubulin interaction appears to include close contacts between taxol's phenyl groups and hydrophobic residues such as Phe, Val, and Leu. Taxol was originally extracted from the slow-growing and endangered Pacific yew tree, but it can also be purified from more renewable sources or chemically synthesized. Taxol is used as an anticancer agent because it blocks cell division and is therefore toxic to rapidly dividing cells such as tumor cells. Keratin is an intermediate filament In addition to actin filaments and microtubules, eukaryotic cells—particularly those in multicellular organisms—contain intermediate filaments. With a diameter of about 100 Å, these fibers are intermediate in thickness to actin filaments and microtubules. Intermediate filaments are exclusively structural proteins. They play no part in cell motility, and unlike actin filaments and microtubules, they have no associated motor proteins. However, they do interact with actin filaments and microtubules via cross-linking proteins. Intermediate filament proteins as a group are much more heterogeneous than the highly conserved actin and tubulin. For example, humans have about 65 intermediate filament genes. The lamins are the intermediate filaments that help form the nuclear lamina in animal cells, a 30–100-Å-thick network inside the nuclear membrane that helps define the nuclear shape and may play a role in DNA replication and transcription. In many cells, intermediate filaments are much more abundant than microfilaments or microtubules and are most prominent in the dead remnants of epidermal cells—that is, in the hard outer layers of the skin—where they may account for 85% of the total protein (Fig. 5-23). The best-known intermediate filament proteins are the keratins, a large group of proteins that include the "soft" keratins, which help define internal body structures, and the "hard" keratins of skin, hair, and claws. The basic structural unit of an intermediate filament is a dimer of ␣ helices that wind around each other—that is, a coiled coil. The amino acid sequence in such a structure consists of seven-residue repeating units in which the first and fourth residues are predominantly nonpolar. In an a helix, these nonpolar residues line up along one side (Fig. 5-24). Because a nonpolar group appears on average every 3.5 Figure 5-24 Arrangement of 3 7 2 residues in a coiled coil. This view down the axis of two seven-residue a helices shows that amino acids at positions 1 and 4 line up on one side of each helix. Nonpolar residues occupying these positions form a hydrophobic strip along the sides of the helices. 5 4 1 6 6 1 4 2 5 7 3 (a) (b) (c) 139 Structural Proteins Figure 5-25 Three views of a coiled coil. These models show a segment of the coiled coil from the protein tropomyosin. (a) Backbone model. (b) Stick model. (c) Spacefilling model. Each a-helical chain contains 100 residues. The nonpolar strips along each helix contact each other, so the two helices wind around each other in a gentle lefthanded coil. [Structure of tropomyosin (pdb 1C1G) determined by F. G. Whitby and G. N. Phillips, Jr.] ? Which nonpolar residues would be most likely to appear at positions 1 and 4? Which would be least likely? residues but there are 3.6 residues per a-helical turn, the strip of nonpolar residues actually winds slightly around the surface of the helix. Two helices whose nonpolar strips contact each other therefore adopt a coiled structure with a left-handed twist (Fig. 5-25). Each intermediate filament subunit contains a stretch of a helix flanked by nonhelical regions at the N- and C-termini. Two of these polypeptides interact in register (parallel and with ends aligned) to form a coiled coil. The dimers then associate in a staggered antiparallel arrangement to form higher-order fibrous structures (Fig. 5-26). The fully assembled intermediate filament may consist of 16 to 32 polypeptides in cross-section. Note that no nucleotides are required for intermediate filament assembly. The N- and C-terminal domains may help align subunits during polymerization and interact with proteins that cross-link intermediate filaments to other cell components. Keratin fibers themselves are crosslinked through disulfide bonds between Cys residues on adjacent chains. An animal hair—for example, sheep's wool or the hair on your head—consists almost entirely of keratin filaments (Fig. 5-27). Hair resists deformation but can Monomer Dimer (coiled coil) Tetramer Octamer Intermediate filament Figure 5-26 Model of an intermediate filament. Pairs of polypeptides form coiled coils. These dimers associate to form tetramers and so on, ultimately producing an intermediate filament composed of 16 to 32 polypeptides in cross-section. Although drawn as straight rods here, the intermediate filament and its component structures are probably all twisted around each other in some way, much like a man-made rope or cable. 140 Ch 5 Protein Function Figure 5-27 Scanning electron micrograph of a human hair. [Tony Brain/ Science Photo Library/Photo Researchers.] be stretched. Tensile stress breaks the hydrogen bonds between carbonyl and amide groups four residues apart in the keratin a helix. The helices can then be pulled until the polypeptides are fully extended. Additional force causes the polypeptide chain to break. If unbroken, the protein can spring back—at least partially—to its original a-helical conformation when the force is removed. This is why a wool sweater stretched out of shape gradually reverts to its former style. Perhaps due to their cable-like construction, intermediate filaments undergo less remodeling than cellular fibers constructed from globular subunits such as actin and tubulin. For example, nuclear lamins disassemble and reassemble only once during the cell cycle, when the cell divides. Keratins, as part of dead cells, remain intact for years. In the innermost layers of animal skin, epidermal cells synthesize large amounts of keratin. As layers of cells move outward and die, their keratin molecules are pushed together to form a strong waterproof coating. Because keratin is so important for the integrity of the epidermis, mutations in keratin genes are linked to certain skin disorders. In diseases such as epidermolysis bullosa simplex (EBS), cells rupture when subject to normal mechanical stress. The result is separation of epidermal layers, which leads to blistering. The most severe cases of the disease arise from mutations in the most highly conserved regions of the keratin molecules, near the ends of the a-helical regions. Collagen is a triple helix Unicellular organisms can get by with just a cytoskeleton, but multicellular animals must have a way to hold their cells together according to some characteristic body plan. Large animals—especially nonaquatic ones—must also support the body's weight. This support is provided by collagen, which is the most abundant animal protein. It plays a major structural role in the extracellular matrix (the material that helps hold cells together), in connective tissue within and between organs, and in bone. Its name was derived from the French word for glue (at a time when glue was derived from animal connective tissue). There are at least 28 types of collagen with different three-dimensional structures and physiological functions. The most familiar is the collagen from animal bones and tendons, which forms thick, ropelike fibers (Fig. 5-28). This type of collagen is a trimeric molecule about 3000 Å long but only 15 Å wide. As in all forms of collagen, the polypeptide chains have an unusual amino acid composition and an unusual conformation. Except in the extreme N- and C-terminal regions of the polypeptides (which are cleaved off once the protein exits the cell), every third amino acid is Gly, and about 30% of the remaining residues are proline and hydroxyproline (Hyp). Hyp residues result from the hydroxylation of Pro residues after the polypeptide has been synthesized, in a reaction that requires ascorbate (vitamin C; Box 5-D). N Proline collagen proteins yield structures with a diameter of 500 to 2000 Å. [J. Gross/ Science Photo Library/Photo Researchers.] C C CH2 C H2 collagen fibers. Thousands of aligned O CH H2C Figure 5-28 Electron micrograph of O N CH H2C CH2 C HO H Hydroxyproline (Hyp) For a stretch of about 1000 residues, each collagen chain consists of repeating triplets, the most common of which is Gly–Pro–Hyp. Gly residues, which have only a hydrogen atom for a side chain, can normally adopt a wide range of secondary structures. However, the imino groups of Pro and Hyp residues (that is, their connected side chains and amino groups) constrain the geometry of the peptide group. The most stable conformation for a polypeptide sequence containing repeating units of Gly –Pro –Hyp is a narrow left-handed helix (Fig. 5-29a). BOX 5-D 141 Structural Proteins BIOCHEMISTRY NOTE Vitamin C Deficiency Causes Scurvy OH O O C O⩾ HO CH2OH H Ascorbate (vitamin C) In the absence of ascorbate (vitamin C), collagen contains too few hydroxyproline residues and hydroxylated lysyl residues, so the resulting collagen fibers are relatively weak. Ascorbate also participates in enzymatic reactions involved in fatty acid breakdown and production of certain hormones. The symptoms of ascorbate deficiency include poor wound healing, loss of teeth, and easy bleeding—all of which can be attributed to abnormal collagen synthesis—as well as lethargy and depression. Historically, ascorbate deficiency, known as scurvy, was common in sailors on long voyages where fresh fruit was unavailable. A remedy, in the form of a daily ration of limes, was discovered in the mid-eighteenth century. Unfortunately, citrus juice, which was also widely administered, proved much less effective in preventing scurvy

because ascorbate is destroyed by heating and by prolonged exposure to air. For this reason, factors such as exercise and good hygiene were also believed to prevent scurvy. Fruit is not the only source of ascorbate. Most animals—with the exception of bats, guinea pigs, and primates—produce ascorbate, so a diet containing fresh meat can supply sufficient ascorbate, which is vital in locations, such as the far North, where fruit is not available. Given the presence of ascorbate in many foods, a true deficiency is rare in adults. Scurvy does still occur, however, as a side effect of general malnutrition or in individuals consuming odd diets. Fortunately, the symptoms of scurvy, which is otherwise fatal, can be easily reversed by administering ascorbate or consuming fresh food. Question: Refer to the structure of ascorbate to explain why it does not accumulate in the body but instead is readily lost via the kidneys. (a) (b) (c) (d) Figure 5-29 Collagen structure. (a) A sequence of repeating Gly– Pro–Hyp residues adopts a secondary structure in which the polypeptide forms a narrow left-handed helix. The residues in this stick model are color-coded: Gly gray, Pro orange, Hyp red. H atoms are not shown. (b) Space-filling model of a single collagen polypeptide. (c) Space-filling model of the triple helix. (d) Backbone trace showing the three polypeptides in different shades of gray. Each polypeptide has a left-handed twist, but the triple helix has a right-handed twist. [Model of collagen (pdb 2CLG) constructed by J. M. Chen.] 142 G G G G G G G Figure 5-30 Cross-section of the collagen triple helix. In this view, looking down the axis of the threechain molecule, each ball represents an amino acid, and the bars represent peptide bonds. Gly residues (which lack side chains and are marked by “G”) are located in the center of the triple helix, whereas the side chains of other residues point outward from the triple helix. ? Compare this axial view of a collagen triple helix to the axial view of a coiled coil (Fig. 5-24). In collagen, three polypeptides wind around each other to form a right-handed triple helix (Fig. 5-29b–d). The chains are parallel but staggered by one residue so that Gly appears at every position along the axis of the triple helix. The Gly residues are all located in the center of the helix, whereas all other residues are on the periphery. A look down the axis of the triple helix shows why Gly— but no other residue—occurs in the center of the helix (Fig. 5-30). The side chain of any other residue would be too large to fit. In fact, replacing Gly with Ala, the next-smallest amino acid, greatly perturbs the structure of the triple helix. The collagen triple helix is stabilized through hydrogen bonding. One set of interactions links the backbone NOH group of each Gly residue to a backbone CPO group in another chain. The geometry of the triple helix prevents the other backbone NOH and CPO groups from forming hydrogen bonds with each other, but they are able to interact with a highly ordered network of water molecules surrounding the triple helix like a sheath. Collagen molecules are covalently cross-linked Trimeric collagen molecules assemble in the endoplasmic reticulum. After they are secreted from the cell, they are trimmed by proteases and align side-to-side and endto-end to form the enormous fibers visible by electron microscopy (see Fig. 5-28). The fibers are strengthened by several kinds of cross-links. Because collagen polypeptides contain almost no cysteine, these links are not disulfide bonds. Instead, the cross-links are covalent bonds between side chains that have been chemically modified following polypeptide synthesis. For example, one kind of cross-link requires the enzymecatalyzed oxidation of two Lys side chains, which then react to form a covalent bond. The number of these and other types of cross-links tends to increase with age, which explains why meat from older animals is tougher than meat from younger animals. C O C (CH2)3 CH CH NH3 ⩽ H3N CH2 Lys NH C CH2 ⩽ (CH2)3 Lys O O (CH2)3 HC CH NH O CH C (CH2)3 NH CH NH C CH NH O (CH2)2 C C CH O CH (CH2)3 CH NH O O O Collagen fibers have tremendous tensile strength. On a per-weight basis, collagen is stronger than steel. Not all types of collagen form thick linear fibers, however. Many nonfibrillar collagens form sheetlike networks of fibers that support layers of cells in tissues. Often, several types of collagen are found together. Not surprisingly, defects in collagen affect a variety of organ systems (Box 5-E). 143 Structural Proteins CONCEPT REVIEW • • • • • • • • • What is the basis of actin filament polarity? How does actin filament remodeling affect the cell’s shape? How does a microtubule differ structurally from an actin filament? Why does normal physiological function require regulation of the assembly and disassembly of cytoskeletal elements? How do the structures of intermediate filaments such as keratin differ from the structures of actin filaments and microtubules? How do two a helices form a coiled coil? How does the structure of collagen differ from the structure of keratin or the structures of fibers with globular subunits? What is the structural importance of Gly residues in collagen? How are collagen fibers stabilized and cross-linked? BOX 5-E CLINICAL CONNECTION Genetic Collagen Diseases Connective tissue, such as cartilage and bone, consists of cells embedded in a matrix containing proteins (mainly collagen) and a space-filling “ground substance” (mostly polysaccharides; see Section 11-3). The polysaccharides, which are highly hydrated, are resilient and return to their original shape after being compressed. The collagen fibers are strong and relatively rigid, resisting tensile (stretching) forces. Together, the polysaccharides and collagen give ligaments (which attach bone to bone) and tendons (which join muscles to bones) the appropriate degree of resistance and flexibility. The connective tissues that surround muscles and organs contain collagen fibers arranged in sheetlike networks with similar physical properties. The regular arrays of collagen fibers (horizontal bands) in tendon allow the tissue to resist tension. Fibroblasts, collagen-producing cells, occupy the spaces between the fibers. Cartilage typically lacks blood vessels. [Mark Nielsen.] In bone, the extracellular matrix is supplemented by minerals, mainly hydroxyapatite, Ca(PO4)3(OH). This form of calcium phosphate forms extensive white crystals that account for up to 50% of the mass of bone. By itself, calcium phosphate is brittle. Yet bone is thousands of times stronger than the crystals because it is a composite structure in which the mineral is interspersed with collagen fibers. This layered arrangement dissipates stresses so that the bone remains strong but can “give” a little under pressure without shattering. Bone tissue also includes passageways for small blood vessels. During development, most of the skeleton takes shape as cartilaginous tissue becomes mineralized, forming hard bone. The repair of a fractured bone follows a similar process in which fibroblasts moving to the injured site synthesize large amounts of collagen, chondroblasts produce cartilage, and osteoblasts gradually replace the cartilage with bony tissue. Mature bone continues to undergo remodeling, which begins with the release of enzymes and acid from cells known as osteoclasts. The enzymes digest collagen and other extracellular matrix components, while the low pH helps dissolve calcium phosphate. Osteoblasts then fill the void with new bone material. Some bones are remodeled faster than others, and the system responds to physical demands so that the bone becomes stronger and thicker when subjected to heavy loads. This is the basis for orthodontics: Teeth are realigned by placing stress on the bone in the tooth sockets for a period long enough for remodeling to take place. The importance of collagen for the structure and function of connective tissues means that irregularities in the collagen protein itself or in the enzymes that process collagen molecules can lead to serious physical abnormalities. Hundreds of collagen-related mutations have been identified. Because most tissues contain more than one type of collagen, the physiological manifestations of collagen mutations are highly variable. Defects in collagen type I (the major form in bones and tendons) cause the congenital disease osteogenesis imperfecta. The primary symptoms of the disease include bone fragility leading to easy fracture, long-bone deformation, and abnormalities of the skin and teeth. X-Ray of a child with a moderately severe case of osteogenesis imperfecta. [ISM/Phototake.] Mutations in collagen type II, a form found in cartilage, lead to osteoarthritis. This genetic disease, which becomes apparent in childhood, is distinct from the osteoarthritis that can develop later in life, often after years of wear and tear on the joints. Defects in the proteins that process collagen extracellularly and help assemble collagen fibers lead to disorders such as dermatosparaxis, which is characterized by extreme skin fragility. Ehlers–Danlos syndrome results from abnormalities in collagen type III, a molecule that is abundant in most tissues but is scarce in skin and bone. Symptoms of this phenotypically variable disorder include easy bruising, thin or elastic skin, and joint hyperextensibility. In one form of the disease, which is accompanied by a high risk for arterial rupture, the molecular defect is a mutation in a collagen type III gene. In another form of the disease, in which individuals often suffer from scoliosis (curvature of the spine), the collagen genes are normal. In these cases, the disease results from a deficiency of lysyl oxidase, the enzyme that modifies Lys residues so that they can participate in collagen cross-links. Ehlers–Danlos syndrome is both rarer and less severe than osteogenesis imperfecta, with many affected individuals surviving to adulthood. Collagen type I, a trimeric molecule, contains two different types of polypeptide chains. Therefore, the severity of the disease depends in part on whether one or two chains in a collagen molecule are affected. Furthermore, the location and nature of the mutation determine whether the abnormal collagen retains some normal function. For example, in one severe form of osteogenesis imperfecta, a 599-base deletion in a collagen gene represents the loss of a large portion of triple helix. The resulting protein is unstable and is degraded intracellularly. Milder cases of osteogenesis imperfecta result from amino acid substitutions, for example, the replacement of Gly by a bulkier residue. Other amino acid changes may slow intracellular processing and excretion of collagen polypeptides, which affects the assembly of collagen fibers. Osteogenesis imperfecta affects about one in 10,000 people. Questions: 1. Explain why a broken bone, which is hard, can heal within a few weeks, whereas a torn ligament (which is softer) may take months to heal. 2. Osteoporosis is the weakening of bones that commonly occurs with aging. Why is exercise recommended as a strategy for preventing fractures in the elderly? 3. Osteogenesis imperfecta is a genetic disease, but most cases result from new mutations rather than ones that are passed from parent to child. Explain. 4. Because networks of collagen fibers lend resilience to the deep (living) layers of skin, some “anti-aging” cosmetics contain purified collagen. Explain why this exogenous (externally supplied) collagen is unlikely to attenuate the skin wrinkling that accompanies aging. 5-3 Motor Proteins KEY CONCEPTS • The motor protein myosin couples the steps of ATP hydrolysis to conformational changes, resulting in muscle contraction. • Kinesin transports cargo by moving processively along a microtubule track. CAN proteins move? An assortment of molecular machinery acts on structural proteins—actin filaments and microtubules—to generate the movements that allow cells to reorganize their contents, change their shape, and even crawl or swim. For example, muscle contraction is accomplished by the motor protein myosin pulling on actin filaments. Eukaryotic cells that move via the wavelike motion of cilia or flagella rely on the motor protein dynein, which acts by bending a bundle of microtubules. Intracellular transport is carried out by motor proteins that move along microfilament and microtubule tracks. In this section, we focus on two well-characterized motor proteins, myosin and kinesin, which provide additional examples of the ways that proteins can move. Myosin has two heads and a long tail There are at least 20 different types of myosin, which is present in nearly all eukaryotic cells. Myosin works with actin to produce movement by transducing chemical energy (in the form of ATP) to mechanical energy. Muscle myosin, with a total 145 Motor Proteins Head Tail Neck Figure 5-31 Structure of muscle myosin. (a) Electron micrograph. [Courtesy John Trinick, University of Leeds.] 1600 Å 165 Å (b) (b) Drawing of a myosin molecule. Myosin’s two globular heads are connected via necks to myosin’s tail, where the polypeptide chains form a coiled coil. ? (a) molecular mass of about 540 kD, consists mostly of two large polypeptides that form two globular heads attached to a long tail (Fig. 5-31). Each head includes a binding site for actin and a binding site for an adenine nucleotide. In the tail region, the two polypeptides twist around each other to form a single rodlike coiled coil (the same structural motif that occurs in intermediate filaments). The “neck” that joins each myosin head to the tail region consists of an a helix about 100 Å long, around which are wrapped two small polypeptides (Fig. 5-32). These so-called light chains help stiffen the neck helix so that it can act as a lever. Each myosin head can interact noncovalently with a subunit in an actin filament, but the two heads act independently, and only one head binds to the actin filament at a given time. In a series of steps that include protein conformational changes and the hydrolysis of ATP, the myosin head releases its bound actin subunit and rebinds another subunit closer to the (1) end of the actin filament. Repetition of this reaction cycle allows myosin to progressively walk along the length of the actin filament. In a muscle cell, hundreds of myosin tails associate to form a thick filament with the head domains sticking out (Fig. 5-33). These heads act as cross-bridges to thin filaments, which each consist of an actin filament and actin-binding proteins that regulate the accessibility of the actin subunits to myosin heads. When a muscle contracts, the multitude of myosin heads individually bind and release actin, like rowers working asynchronously, which causes the thin and thick filaments to slide past each other (Fig. 5-34). Because of the arrangement of filaments in the muscle cell, the action of myosin on actin results in an overall shortening of the muscle. This phenomenon is commonly called contraction, but the muscle does not undergo any compression In what part of the protein are large hydrophobic residues most likely to occur? Neck ATP-binding pocket Head Actinbinding site Figure 5-32 Myosin head and neck region. The myosin neck forms a molecular lever between the head domain and the tail. Two light chains (lighter shades of purple) help stabilize the a-helical neck. The actin-binding site is at the far end of the myosin head. ATP binds in a cleft near the middle of the head. Only the alpha carbons of the light chains are visible in this model. [Structure of chicken myosin (pdb 2MYS) determined by I. Rayment and H. M. Holden.] Figure 5-33 Electron micrograph of a thick filament. The heads of many myosin molecules project laterally from the thick rod formed by the aligned myosin tails. [From Trinick, J., and Elliott, A., J. Mol. Biol. 131, 15 (1977).] 146 Ch 5 Protein Function Thin filament (actin) Thick filament (myosin) Relaxed Contracted Figure 5-34 Movement of thin and thick filaments during muscle contraction. and its volume remains constant—it actually becomes thicker around the middle. A shortening on the order of 20% of a muscle’s length is typical; 40% is extreme. Myosin operates through a lever mechanism How does myosin work? The key to its mechanism is the hydrolysis of the ATP that is bound to the myosin head. Although the ATP-binding site is about 35 Å away from the actin-binding site, the conversion of ATP to ADP 1 Pi triggers conformational changes in the myosin head that are communicated to the actin-binding site as well as to the lever (the neck region). The chemical reaction of ATP hydrolysis thereby drives the physical movement of myosin along an actin filament. In other words, the free energy released by hydrolysis of ATP is transformed into mechanical work. The four steps of the myosin–actin reaction sequence are shown in Figure 5-35. Note how each event at the nucleotide-binding site correlates with a conformational change related to either actin binding or bending of the lever. An a helix makes an ideal lever because it can be quite long. It is also relatively incompressible, so it can pull the coiled-coil myosin tail along with it. Altogether, the lever swings by about 708 relative to the myosin head. The return of the lever to its original conformation (step 4 of the reaction cycle) is the force-generating step. When adjusted for the difference in mass, the myosin–actin system has a power output comparable to a typical automobile. Each cycle of ATP hydrolysis moves the myosin head by an estimated 50–100 Å. Since individual actin subunits are spaced about 55 Å apart along the thin filament, the myosin head advances by at least one actin subunit per reaction cycle. Because the reaction cycle involves several steps, some of which are essentially irreversible (such as ATP S ADP 1 Pi), the entire cycle is unidirectional. Myosin operates in many cells, not just in muscles. For example, myosin works with actin during cytokinesis (the splitting of the cell into two halves following mitosis), and some myosin proteins use their motor activity to transport certain cell components along microfilament tracks. Myosin molecules may also act as tension rods to cross-link the microfilaments of the cytoskeleton. This is one reason why mutations in myosin in the sensory cells of the ear cause deafness and other abnormalities (Box 5-F). Kinesin is a microtubule-associated motor protein Many cells also contain motor proteins, such as kinesin, that move along microtubule tracks. There are several different types of kinesins; we will describe the prototypical one, also known as conventional kinesin. Myosin Neck Actin Head 147 Motor Proteins 1. The reaction sequence begins with a myosin head bound to an actin subunit of the thin filament. ATP binding alters the configuration of the myosin head so that it releases actin. ATP ATP 2. The rapid hydrolysis of ATP to ADP + Pi triggers a conformational change that rotates the myosin lever and increases the affinity of myosin for actin. ADP Pi 3. Myosin binds to an actin subunit farther along the thin filament. ADP Pi Pi ADP 4. Binding to actin causes Pi and then ADP to be released. As these reaction products exit, the myosin lever returns to its original position. This causes the thin filament to move relative to the thick filament (the power stroke). ATP replaces the lost ADP to repeat the reaction cycle. Figure 5-35 The myosin–actin reaction cycle. For simplicity, only one myosin head is shown. See Animated Figure. Mechanism of force generation in muscle. ? BOX 5-F BIOCHEMISTRY NOTE Myosin Mutations and Deafness Inside the cochlea, the spiral-shaped organ of the inner ear, are thousands of hair cells, each of which is topped with a bundle of bristles known as stereocilia. [P. Motta/Science Photo Library/Photo Researchers.] Each stereocilium contains several hundred cross-linked actin filaments and is therefore extremely rigid, except at its base, where there are fewer actin filaments. Sound waves deflect the stereocilia at the base, initiating an electrical signal that is transmitted to the brain. (continued on the next page) Write an equation that describes the reaction catalyzed by myosin. Myosin molecules probably help control the tension inside each stereocilium, so the ratcheting activity of the myosin motors along the actin filaments may adjust the sensitivity of the hair cells to different degrees of stimulus. Other myosin molecules whose tails bind certain cell constituents may use their motor activity to redistribute these substances along the length of the actin filaments. Abnormalities in any of these proteins could interfere with normal hearing. About half of all cases of deafness have a genetic basis, and over 100 different genes have been linked to deafness. One of these codes for myosin type VIIa, which is considered to be an “unconventional” myosin because it differs somewhat from the “conventional” myosin of skeletal muscle (also known as myosin type II). Gene-sequencing studies indicate that myosin VIIa has 2215 residues and forms a dimer with two heads and a long tail. Its head domains probably operate by the same mechanism as muscle myosin, converting the chemical energy of ATP into mechanical energy for movement relative to an actin filament. Over a hundred mutations in the myosin VIIa gene have been identified, including premature stop codons, amino acid substitutions, and deletions—all of which compromise the protein’s function. Such mutations are responsible for many cases of Usher syndrome, the most common form of deaf-blindness in the United States. Usher syndrome is characterized by profound hearing loss, retinitis pigmentosa (which leads to blindness), and sometimes vestibular (balance) problems. The congenital deafness of Usher syndrome results from the failure of the cochlear hair cells to develop properly. The unresponsiveness of the stereocilia to sound waves probably also accounts for their inability to respond normally to the movement of fluid in the inner ear, which is necessary for maintaining balance. Abnormal myosin also plays a role in the blindness that often develops in individuals with Usher syndrome, usually by the second or third decade. The intracellular transport function of myosin VIIa is responsible for distributing bundles of pigment in the retina. In retinitis pigmentosa, retinal neurons gradually lose their ability to transmit signals in response to light; in advanced stages of the disease, pigment actually becomes clumped on the retina. 148 Ch 5 Protein Function Question: Not all mutations are associated with a loss of function. Explain how a myosin mutation that increases the rate of ADP release could lead to deafness. Kinesin, like myosin, is a relatively large protein (with a molecular mass of 380 kD) and has two large globular heads and a coiled-coil tail domain (Fig. 5-36). Each 100-Å-long head consists of an eight-stranded b sheet flanked by three a helices on each side and includes a tubulin-binding site and a nucleotide-binding site. The light chains, situated at the opposite end of the protein, bind to proteins in the (a) (b) Tail Heads Light chains Figure 5-36 Structure of kinesin. (a) Diagram of the molecule. (b) Model of the head and neck region. Each globular head, which contains a helices and b sheets, connects to an a helix that winds around its counterpart to form a coiled coil. Two light chains at the end of the coiled-coil tail can interact with a membranous vesicle, the “cargo.” [Structure of rat kinesin (pdb 3KIN) determined by F. Kozielski, S. Sack, A. Marx, M. Thormahlen, E. Schonbrunn, V Biou, A. Thompson, E.-M. Mandelkow, and E. Mandelkow.] ? Compare the structure of kinesin to the structure of myosin (Fig. 5-31). membrane shell of a vesicle. The vesicle and its contents become kinesin’s cargo. Kinesin moves its cargo toward the (1) end of a microtubule by stepping along the length of a single protofilament. Other microtubule-associated motor proteins appear to use a similar mechanism but move toward the (2) end of the microtubule. The motor activity of kinesin requires the chemical energy of ATP hydrolysis. However, kinesin cannot follow the myosin lever mechanism because its head domains are not rigidly fixed to its neck regions. In kinesin, a relatively flexible polypeptide segment joins each head to a helix that eventually becomes part of the coiled-coil tail (see Fig. 5-36b). (Recall that in myosin, the lever is a long a helix that extends from the head to the coiled-coil region and is stiffened by the two light chains; see Fig. 5-32.) Nevertheless, the relative flexibility of kinesin’s neck is critical for its function. Kinesin’s two heads are not independent but work in a coordinated fashion so that the two heads alternately bind to successive b-tubulin subunits along a protofilament, as if walking. Conformational changes elicited by ATP binding and hydrolysis are relayed to other regions of the molecule (Fig. 5-37). This transforms the free energy of ATP into the mechanical movement of kinesin. Each ATP-binding event Figure 5-37 The kinesin reaction Cargo cycle. The cycle begins with one kinesin head bound to a tubulin subunit in a protofilament of a microtubule. The trailing head has ADP in its nucleotidebinding site. For clarity, the relative size of the neck region is exaggerated. ADP (+) β α Protofilament 1. ATP binding to the leading head (orange) induces a conformational change in which the neck docks against the head. This movement swings the trailing head (yellow) forward by 180° toward the (+) end of the microtubule. This is the force-generating step. ATP ADP ATP α β ADP 2. The new leading head (yellow) quickly binds to a tubulin subunit and releases its ADP. This step moves kinesin’s cargo forward along the protofilament. ATP α β Pi 3. In the trailing head (orange), ATP is hydrolyzed to ADP + Pi . The Pi diffuses away, and the trailing head begins to detach from the microtubule. ADP α β 149 Motor Proteins 4. ATP binds to the leading head to repeat the reaction cycle. ? What prevents kinesin from moving backward? 150 Ch 5 Protein Function yanks the trailing head forward by about 160 Å, so the net movement of the attached cargo is about 80 Å, or the length of a tubulin dimer. Kinesin is a processive motor Figure 5-38 Electron micrograph of neurons. Microtubule-associated motor proteins move cargo between the cell body and the ends of the axon and other cell processes. [CNRI/Science Photo Library/Photo Researchers.] As in the myosin–actin system (see Fig. 5-35), the kinesin–tubulin reaction cycle proceeds in only one direction. Although most molecular movement is associated with ATP binding, ATP hydrolysis is a necessary part of the reaction cycle. The slowest step of the reaction cycle shown in Figure 5-37 is the dissociation of the trailing kinesin head from the microtubule. ATP binding to the leading head may help promote trailing-head release as part of the forward-swing step. Because the free head quickly rebinds tubulin, the kinesin heads spend most of their time bound to the microtubule track. One consequence of kinesin’s almost constant hold on a microtubule is that many—perhaps 100 or more—cycles of ATP hydrolysis and kinesin advancement can occur before the motor dissociates from its microtubule track. Kinesin is therefore said to have high processivity. A motor protein such as myosin, which dissociates from an actin filament after a single stroke, is not processive. In a muscle cell, low processivity is permitted because the many myosin–actin interactions occur more or less simultaneously to cause the thin and thick filaments to slide past each other (see Fig. 5-34). High processivity is advantageous for a transport engine such as kinesin, because its cargo (which is relatively large and bulky) can be moved long distances without being lost. Consider the need for efficient transport in a neuron. Neurotransmitters and membrane components are synthesized in the cell body, where ribosomes are located, but must be moved to the end of the axon, which may be several meters long in some cases (Fig. 5-38). CONCEPT REVIEW • Describe the steps of the myosin–actin reaction cycle. • Which events at the nucleotide-binding site correlate with protein conformational changes? • Compare and contrast the myosin–actin and kinesin–tubulin reaction cycles. • Why is kinesin a processive motor? SUMMARY 5-1 Myoglobin and Hemoglobin: Oxygen-Binding Proteins • Myoglobin contains a heme prosthetic group that reversibly binds oxygen. The amount of O2 bound depends on the O2 concentration and on myoglobin’s affinity for oxygen. • Hemoglobin’s a and b chains are homologous to myoglobin, indicating a common evolutionary origin. • O2 binds cooperatively to hemoglobin with low affinity, so hemoglobin can efficiently bind O2 in the lungs and deliver it to myoglobin in the tissues. • Hemoglobin is an allosteric protein whose four subunits alternate between the T (deoxy) and R (oxy) conformations in response to O2 binding to the heme groups. The deoxy conformation is favored by low pH (the Bohr effect) and by the presence of BPG. 5-2 Structural Proteins • The microfilament elements of a cell’s cytoskeleton are built from ATP-binding globular actin subunits that polymerize as a double chain. Polymerization is reversible, so actin filaments undergo growth and regression. Their dynamics may be modified by proteins that mediate filament capping, branching, and severing. • GTP-binding tubulin dimers polymerize to form a hollow microtubule. Polymerization is more rapid at one end, and the microtubule can disassemble rapidly by fraying. Drugs that affect microtubule dynamics interfere with cell division. • The intermediate filament keratin contains two long a-helical chains that coil around each other so that their hydrophobic residues are in contact. Keratin filaments associate and are crosslinked to form semipermanent structures. • Collagen polypeptides contain a large amount of proline and hydroxyproline and include a Gly residue at every third position. Each chain forms a narrow left-handed helix, and three chains coil around each other to form a right-handed triple helix with Gly residues at its center. Covalent cross-links strengthen collagen fibers. 5-3 Motor Proteins • Myosin, a protein with a coiled-coil tail and two globular heads, interacts with actin filaments to perform mechanical work. ATPdriven conformational changes allow the myosin head to bind, release, and rebind actin. This mechanism, in which myosin acts as a lever, is the basis of muscle contraction. • The motor protein kinesin has two globular heads connected by flexible necks to a coiled-coil tail. Kinesin transports vesicular cargo along the length of a microtubule by a processive stepping mechanism that is driven by conformational changes triggered by ATP binding and hydrolysis. 151 GLOSSARY TERMS cytoskeleton motor protein heme prosthetic group Y pO2 saturation p50 globin homologous proteins invariant residue conservative substitution variable residue cooperative binding deoxyhemoglobin BIOINFORMATICS oxyhemoglobin T state R state allosteric protein ligand anemia thalassemia Bohr effect microfilament intermediate filament microtubule F-actin G-actin (2) end (1) end PROJECT 4 treadmilling protofilament gout coiled coil imino group triple helix osteogenesis imperfecta Ehlers–Danlos syndrome thick filament thin filament stereocilia Usher syndrome cytokinesis vesicle processivity Learn to use database tools to compare and align protein sequences and construct phylogenetic trees. USING DATABASES TO COMPARE AND INDENTIFY RELATED PROTEIN SEQUENCES PROBLEMS 5-1 Myoglobin and Hemoglobin: Oxygen-Binding Proteins 1. Explain why globin alone or heme alone is not effective as an oxygen carrier. 2. Myoglobin transfers oxygen obtained from hemoglobin to mitochondrial proteins involved in the catabolism of metabolic fuels to produce energy for the muscle cell. Using what you have learned in this chapter about myoglobin and hemoglobin, what can you conclude about the structures of these mitochondrial proteins? 1.00 0.75 Y 0.50 Tuna Bonito Mackerel 0.25 3. Myoglobin most effectively facilitates oxygen diffusion through muscle cells when the intracellular oxygen partial pressure is comparable to the p50 value of myoglobin. Explain why. 4. Heart and muscle cells, where myoglobin resides, maintain an intracellular oxygen partial pressure of about 2.5 torr. Explain why a small change in oxygen partial pressure of 1 torr in either direction results in a dramatic change in myoglobin oxygen binding. 5. Oxygenated myoglobin is bright red, whereas deoxygenated myoglobin is a purplish color. Myoglobin in which the Fe21 has been oxidized to Fe31 is referred to as met-myoglobin and appears brown. The sixth ligand in met-myoglobin is water rather than oxygen. (a) When a fresh, uncooked beef roast is sliced in two, the surface of the meat first appears purplish but then turns red. Explain why. (b) Why does meat turn brown when it is cooked? (c) Why might a retailer choose to package meat using oxygenpermeable plastic wrap instead of a vacuum-sealed package? 6. Myoglobins from three species of fish—tuna, bonito, and mackerel—were purified, and their oxygen binding affinities were measured at 208C. The results are shown in the figure above right. (a) Use the graph to estimate the p50 values for each species. (b) Which species of fish has the highest oxygen-binding affinity at the experimental temperature? The least? 2 4 6 8 10 p O2 (torr) 7. The oxygen-binding site in myoglobin, shown in Figure 5-2 and in the figure below, resides in a heme pocket in which a valine residue provides one of the pocket boundaries. How is oxygen binding affected when the Val residue is mutated to an Ile? His E7 N N H O H 3C O CH H3C Val 68 N HN His F8 8. Refer to the figure in Problem 7. How would oxygen-binding affinity be affected if the Val residue was mutated to a Ser? (Hint: Is it more advantageous for oxygen binding for the pocket to be polar or nonpolar?) 152 9. Hexacoordinate Fe(II) in heme is bright red. Pentacoordinate Fe(II) is blue. Explain how these electronic changes account for the different colors of arterial (scarlet) and venous (purple) blood. 10. Carbon monoxide, CO, binds to hemoglobin about 250 times as tightly as oxygen. The resulting complex is a red color that is much brighter than that observed when oxygen is bound. What symptoms would you expect to observe in a patient with CO poisoning? 11. Below is a list of the first 10 residues of the B helix in myoglobin from different organisms. Based on this information, which positions (a) cannot tolerate substitution, (b) can tolerate conservative substitution, and (c) are highly variable? Position Human Chicken Alligator 1 2 3 4 5 6 7 8 9 10 D I P G H G Q E V L D I A G H G H E V L Turtle Tuna Carp D L S A H G Q E V I D Y T T M G G L V L D F E G T G G E V L K L P E H G H E V I 12. Invariant residues are those that are essential for the structure and/or function of the protein and cannot be replaced by other residues. Which amino acid residue in the globin chain is most likely to be invariant? Explain. 13. Hemoglobin is 50% saturated with oxygen when pO2 5 26 torr. If hemoglobin exhibited hyperbolic binding (as myoglobin does) with 50% saturation at 26 torr, what would be the fractional saturation when pO2 5 30 torr and 100 torr? What does this tell you about the physiological importance of hemoglobin’s sigmoidal oxygen-binding curve? 14. But hemoglobin does not exhibit hyperbolic binding, as we assumed in Problem 13; its oxygen-binding curve has a sigmoidal shape. The equation for hemoglobin’s sigmoidal oxygen dissociation curve is modified from Equation 5-4: Y5 (p O2) n (p50) n 1 (p O2) n The quantity n is known as the Hill coefficient, and its value increases with the degree of cooperation of the ligand (in this case, oxygen) binding. For hemoglobin, the value of n is approximately equal to 3. We can use the equation above to calculate the fractional saturation of hemoglobin. (This was first measured in 1904 by Christian Bohr and his colleagues, who measured oxygen binding in dog blood as a function of oxygen partial pressure at different concentrations of carbon dioxide.) At a CO2 partial pressure of 5 torr, the p50 value for hemoglobin is 15 torr. What is the fractional saturation when pO2 5 25 torr, a typical venous oxygen partial pressure? What is the fractional saturation when pO2 5 120 torr, a typical oxygen partial pressure in the lungs? 15. Calculate the fractional saturation of oxygen binding to hemoglobin measured by Bohr and his colleagues for pO2 5 25 torr and 120 torr (see Problem 14) when the CO2 partial pressure is 80 torr. Under these conditions, the p50 value for hemoglobin is 40 torr. 16. Compare your answers to Problems 14 and 15. How much oxygen is delivered to tissues when the pCO2 is 5 torr? How does this compare to the amount of oxygen delivered to tissues when the pCO2 is 40 torr? What role does CO2 play in assisting oxygen delivery to tissues? 17. The grelag goose lives year-round in the Indian plains, while its close relative, the bar-headed goose, spends the summer months in the Tibetan lake region. The p50 value for bar-headed goose hemoglobin is less than the p50 for grelag goose hemoglobin. What is the significance of this adaptation? 18. Drinking a few drops of a preparation called “vitamin O,” which consists of oxygen and sodium chloride dissolved in water, purportedly increases the concentration of oxygen in the body. (a) Use your knowledge of oxygen transport to evaluate this claim. (b) Would vitamin O be more or less effective if it was infused directly into the bloodstream? 19. Highly active muscle generates lactic acid by respiration so fast that the blood passing through the muscle actually experiences a drop in pH from about 7.4 to about 7.2. Under these conditions, hemoglobin releases about 10% more O2 than it does at pH 7.4. Explain. 20. Plasmodium falciparum, the protozoan that causes malaria, slightly decreases the pH of the red blood cells it infects. Invoke the Bohr effect to explain why Plasmodium-infected cells are more likely to undergo sickling in individuals with the Hb S variant. 21. About two dozen histidine residues in hemoglobin are involved in binding the protons produced by cellular metabolism. In this manner, hemoglobin contributes to buffering in the blood, and the imidazole groups able to bind and release protons contribute to the Bohr effect. One important contributor to the Bohr effect is His 146 on the b chain of hemoglobin, whose side chain is in close proximity to the side chain of Asp 94 in the deoxy form of hemoglobin but not the oxy form. (a) What kind of interaction occurs between Asp 94 and His 146 in deoxyhemoglobin? (b) The proximity of Asp 94 alters the pK value of the imidazole ring of His. In what way? 22. Refer to the data given in Problems 14 and 15. Compare the p50 values for hemoglobin when the pCO2 is 5 torr and 80 torr. What is the significance of the different p50 values? 23. Lampreys are among the world’s most primitive vertebrates and are therefore interesting organisms to study. It has been shown that lamprey hemoglobin forms a tetramer when deoxygenated, but when the hemoglobin becomes oxygenated, the tetramers dissociate into monomers. The binding of oxygen to lamprey hemoglobin is influenced by pH; as in human hemoglobin, the deoxygenated form is favored when the pH decreases. Glutamate residues on the surface of the monomers play an important role. How might glutamate residues influence the equilibrium between the monomer and tetramer form with changes in pH? 24. While most humans are able to hold their breath for only a minute or two, crocodiles can stay submerged under water for an hour or longer. This adaptation allows the crocodiles to kill their prey by drowning. In 1995, Nagai and colleagues suggested that bicarbonate (HCO2 3) binds to crocodile hemoglobin to promote oxygen dissociation in a manner similar to BPG in humans and, in so doing, promotes delivery of a large fraction of bound oxygen to tissues. Information gathered in these experiments may allow scientists to design effective blood replacements. 153 (a) Consider the hypothesis that bicarbonate serves as an allosteric modulator of oxygen binding to hemoglobin in crocodiles. What is the source of HCO2 3 in crocodile tissues? (b) Draw oxygen-binding curves for crocodile hemoglobin in the presence and absence of bicarbonate. Which conditions increase the p50 value for crocodile hemoglobin? What does this tell you about the oxygen-binding affinity of hemoglobin under those conditions? (c) The investigators found that the bicarbonate binding site on crocodile hemoglobin was located at the a1b2 subunit interface, where the two subunits slide with respect to each other during the oxy 4 deoxy transition. Based on their results, the researchers modeled a stereochemically plausible binding site that included the phenolate anions of Tyr 41b and Tyr 42a and the ε-amino group of Lys 38b. What interactions might form between these amino acid side chains and bicarbonate? (d) Fish use ATP or GTP as allosteric effectors to encourage hemoglobin to release its bound oxygen. What structural characteristics do all of these different allosteric effectors have in common and how would the molecules bind to hemoglobin? 25. In the developing fetus, fetal hemoglobin (Hb F) is synthesized beginning at the third month of gestation and continuing until birth. After the baby is born, the concentration of Hb F declines and is replaced with adult hemoglobin (Hb A) as synthesis of the g chain declines and synthesis of the b chain increases. By the time the baby is six months old, 98% of its hemoglobin is Hb A. (a) In the graph below, which curve represents fetal hemoglobin? 1.0 Hb Great Lakes 0.8 Hb A 0.6 Y 0.4 0.2 0 10 20 30 40 50 60 p O2 (torr) (a) Compare the shapes of the curves for the two hemoglobins and comment on their significance. (b) Which hemoglobin has a higher affinity for oxygen when pO2 5 20 torr? (c) Which hemoglobin has a higher affinity for oxygen when pO2 5 75 torr? (d) Which hemoglobin is most efficient at delivering oxygen from arterial blood (pO2 5 75 torr) to active muscle (pO2 5 20 torr)? 28. Rank the stability of the following mutant hemoglobins: Hb Hammersmith b42Phe S Ser Hb Bucuresti b42Phe S Leu Hb Sendagi b42Phe S Val Hb Bruxelles b42Phe S 0 (Phe is deleted) 29. People who live at or near sea level undergo high-altitude acclimatization when they go up to moderately high altitudes of about 5000 m. (a) The low concentration of oxygen at these altitudes initially induces hypoxia, which results in hyperventilation. Explain why this occurs. What happens to blood pH during hyperventilation? (b) After a period of weeks, the alveolar pCO2 decreases and the 2,3-bisphosphoglycerate concentration increases. Explain these observations. A Y B p O2 (b) Why does Hb F have a higher oxygen affinity than Hb A? Why does this provide an advantage to the developing fetus? 26. The drug hydroxyurea can be used to treat sickle cell anemia, although it is not used often because of undesirable side effects. Hydroxyurea is thought to function by stimulating the afflicted person’s synthesis of fetal hemoglobin. In a clinical study, patients who took hydroxyurea showed a 50% reduction in frequency of hospital admissions for severe pain, and there was also a decrease in the frequency of fever and abnormal chest X-rays. Why would this drug alleviate the symptoms of sickle cell anemia? 27. The oxygen-binding curves for normal hemoglobin (Hb A) and a mutant hemoglobin (Hb Great Lakes) are shown in the figure. 30. Animals indigenous to high altitudes, such as yaks, llamas, and alpacas, have adapted to the low oxygen concentrations at high altitudes because they have high-affinity hemoglobins, achieved through a combination of globin-chain mutations and persistence of fetal hemoglobin (i.e., fetal hemoglobin is synthesized by mature animals). This is a true evolutionary adaptation and is different from the high-altitude acclimatization described in Problem 29. (a) Compare the p50 values of the high-altitude animals with the p50 values of their low-altitude counterparts. (b) How does the continued synthesis of fetal hemoglobin help these animals survive at high altitudes? 31. A patient with a high-affinity hemoglobin mutant was the subject of a paper published in 1970s. The subject’s a chain was normal, but a mutation in the b chain occurred that resulted in the replacement of a lysine by an asparagine at position 144. (The normal C-terminal sequence of the b chain is –Lys 144–Tyr 154 145–His 146–COO2). The C-terminal portion of the b chain, which resides in the central cavity of the hemoglobin molecule, normally forms interactions that affect the position of the F helix that contains His F8. How might the Lys S Asn substitution result in a higher-affinity hemoglobin? 32. Patients with the types of mutations described in Problem 31 are able to undergo high-altitude acclimatization (described in Problem 29) much more readily than “lowlanders” with normal hemoglobin. Explain why. 5-2 Structural Proteins 33. Compare and contrast globular and fibrous proteins. 34. Of the various proteins highlighted in this chapter (actin, tubulin, keratin, collagen, myosin, and kinesin), (a) which are exclusively structural (not involved in cell shape changes); (b) which are considered to be motor proteins; (c) which are not motor proteins but can undergo structural changes; and (d) which contain nucleotide-binding sites? 35. Explain why microfilaments and microtubules are polar whereas intermediate filaments are not. 36. In order to obtain crystals suitable for X-ray crystallography, G-actin was first mixed with another protein. (a) Why was this step necessary? (b) What additional information was required to solve the structure of G-actin? 37. Phalloidin, a peptide isolated from a poisonous mushroom, binds to F-actin but not G-actin. How does the addition of phalloidin affect cell motility? 38. Phalloidin (see Problem 37) is used as an imaging tool by covalently attaching a fluorescent tag to it. What structures can be visualized in cells treated with fluorescently labeled phalloidin? What is not visible? 39. How could a microtubule-binding protein distinguish a rapidly growing microtubule from one that was growing more slowly? 40. Why would fraying be a faster mechanism for microtubule disassembly than dissociation of tubulin dimers? 41. Microtubule ends with GTP bound tend to be straight, whereas microtubule ends with GDP bound are curved and tend to fray. An experiment was carried out in which monomers of b-tubulin were allowed to polymerize in the presence of a nonhydrolyzable analog of GTP called guanylyl-(ab)-methylenediphosphonate (GMP CPP). Compare the stability of this polymer with one polymerized in the presence of GTP. O N ⩾O O O O P CH2 P O P O⩾ O⩾ GMP . CPP O⩾ N N NH2 O O H2C H NH H H H OH OH 42. Explain why colchicine (which promotes microtubule depolymerization) and paclitaxel (taxol, which prevents depolymerization) both prevent cell division. 43. A recent review listed literally dozens of compounds that are being investigated as possible therapeutic agents for treating cancer. Tubulin was the target for all of the compounds. Why is tubulin a good target for anticancer drugs? 44. The three-dimensional structure of paclitaxel bound to the tubulin dimer has been solved. Paclitaxel binds to a pocket in b-tubulin on the inner surface of the microtubule and causes a conformational change that counteracts the effects of GTP hydrolysis. How does the binding of paclitaxel affect the stability of the microtubule? 45. Gout is a disease characterized by the deposition of sodium urate crystals in the synovial (joint) fluid. The crystals are ingested by neutrophils, which are white blood cells that circulate in the bloodstream before they crawl through tissues to reach sites of injury. Upon ingestion of the crystals, a series of biochemical reactions lead to a release of mediators from the neutrophils that cause inflammation, resulting in joint pain. Colchicine is used as a drug to treat gout because colchicine can inhibit neutrophil mobility. How does colchicine accomplish this? 46. The microtubules involved in cell division are less stable than microtubules found in axon extensions of nerve cells. Why is this the case? 47. Vinblastine, a compound in the periwinkle, has been shown to affect microtubule assembly by stabilizing the (1) ends and destabilizing the (2) ends of microtubules. How does vinblastine affect the formation of the mitotic spindle during mitosis? 48. Why would you expect vinblastine (Problem 47) to have a more dramatic effect on rapidly dividing cells such as cancer cells? 49. Hydrophobic residues usually appear at the first and fourth positions in the seven-residue repeats of polypeptides that form coiled coils. (a) Why do polar or charged residues usually appear in the remaining five positions? (b) Why is the sequence Ile–Gln–Glu–Val–Glu–Arg–Asp more likely than the sequence Trp–Gln–Glu– Tyr–Glu–Arg–Asp to appear in a coiled coil? 50. Globular proteins are typically constructed from several layers of secondary structure, with a hydrophobic core and a hydrophilic surface. Is this true for a fibrous protein such as keratin? 51. Straight hair can be curled and curly hair can be straightened by exposing wet hair to a reducing agent, repositioning the hair with rollers, then exposing the hair to an oxidizing agent. Explain how this procedure alters the shape of the hair. 52. “Hard” keratins such as those found in hair, horn, and nails have a high sulfur content, whereas “soft” keratins found in the skin have a lower sulfur content. Explain the structural basis for this observation. 53. Describe the primary, secondary, tertiary, and quaternary structures of (a) actin and (b) collagen. Why is it difficult to assign the four structural categories to these proteins? 54. A nutrition website advocates the consumption of collagenrich foods such as chicken skin, shark fins, and pigs’ feet as part of a “wrinkle-free” diet. Is the consumption of foods rich in collagen required for proper collagen biosynthesis? Do you think this diet is likely to decrease the appearance of wrinkles? 55. The highly pathogenic bacterium Clostridium perfringens causes gangrene, a disease that results in the destruction of animal tissue. The bacterium secretes two types of collagenases that cleave 155 the peptide bonds linking Gly and Pro residues. A biochemical company sells a partially purified preparation of these collagenases to investigators who are interested in culturing cells derived from bone, cartilage, muscle, or endothelial tissue. How does this enzyme preparation assist the investigator in obtaining cells for culturing? 56. An enzyme found in tadpoles cleaves the bond indicated in the figure below. What kind of enzyme is this, and what is its role in the metamorphosis of the tadpole to the adult frog? Gly Pro Gln Gly Ile Ala 57. Members of the collagen family are often characterized by their melting temperature (Tm). The collagen is subjected to increasing temperature and monitored for structural changes. At high temperatures, the intermolecular forces holding the three chains together in the triple helix are disrupted and the chains come apart, resulting in denatured collagen, or gelatin (a substance that is liquid at high temperatures but solidifies upon cooling). Because the intermolecular forces holding the collagen strands together are cooperative, the collagen triple helix tends to come apart all at once. The Tm is defined as that temperature at which the collagen is half-denatured and is used as a criterion for collagen stability. Tm measurements for collagen from two different species are presented in the table below. One is from rat; the other is from sea urchin. (a) Assign each collagen to the proper organism, and (b) explain the correlation between imino acid content and Tm. Collagen A B Number of Pro per 1000 residues Tm (8C) 48.5 68.5 81.3 111 27.0 38.5 58. A series of collagen-like peptides, each containing 30 amino acid residues, were synthesized in order to study the important interactions among the three chains in the triple helix. Peptide 1 is (Pro–Hyp–Gly)10, peptide 2 is (Pro–Hyp–Gly)4–Glu–Lys–Gly– (Pro–Hyp–Gly)5, and peptide 3 is Gly–Lys–Hyp–Gly–Glu–Hyp– Gly–Pro–Lys–Gly–Asp–Ala–(Gly–Ala–Hyp)2–(Gly–Pro–Hyp)4. Their structural properties and melting temperatures (see Problem 57) are summarized in the table below. Note that “imino acid content” refers to the combined content of Pro and Hyp. Forms trimer Imino acid content 1 Yes 67% 2 Yes 60% 3 Yes 30% Tm (8C) pH 5 1 pH 5 7 pH 5 11 pH 5 1 pH 5 7 pH 5 1 pH 5 7 pH 5 7 pH 5 13 59. The deep-sea hydrothermal vent worm Riftia pachyptila lives under extreme conditions of high temperature, low oxygen content, and drastic temperature changes. The worm has a thick collagencontaining cuticle that protects it from its harsh environment. Recently the structure of this collagen was investigated. Sequence analyses indicated that the collagen had the customary Gly–X–Y triplet but that hydroxyproline occurred only in the X position and that Y was often a glycosylated threonine (a galactose sugar residue was covalently attached via a condensation reaction between a hydroxyl group on the galactose and the hydroxyl group of the threonine side chain). Experiments were carried out using synthetic peptides in order to evaluate the stability of each. Results are shown in the table. OH OH O H H H HO 61 58 60 44 46 49 18 26.5 19 (a) Rank the stability of the three collagen-like peptides. What is the reason for the observed stability? (b) Compare the Tm values of peptide 3 at the various pH values. Why does peptide 3 have a maximum Tm value at pH 5 7? What interactions are primarily responsible? O ⩽H N 3 CH O CH OH H H C O⩾ CH3 Gal-Thr Synthetic peptide Number of Hyp per 1000 residues Peptide (c) Compare the Tm values of peptide 1 at the various pH values. Why is there less of a difference of Tm values at different pH values for peptide 1 than for peptide 3? (d) Consider the answers to the above questions. Which factor is more important in stabilizing the structure of the collagen-like peptides: ion pairing or imino acid content? (Pro–Pro–Gly)10 (Pro–Hyp–Gly)10 (Gly–Pro–Thr)10 (Gly–Pro–Thr(Gal))10 Forms a triple helix? Tm (8C) Yes Yes No Yes 41 60 N/A 41 (a) Compare the melting temperatures of (Pro–Pro–Gly)10 and (Pro–Hyp–Gly)10. What is the structural basis for the difference? (b) Compare the melting temperature of (Pro–Pro–Gly)10 and (Gly–Pro–Thr(Gal))10 and provide an explanation. (c) Why was (Gly–Pro–Thr)10 included by the investigators? 60. Several investigators have sought to explain why hydroxyproline-containing collagen molecules have increased stability. Some suggested that the hydroxyl group on the pyrrolidine ring of Hyp participates in a network of hydrogen bonds with “bridging” water molecules. However, this has been called into question by others, who have noted that triple helices made of (Pro–Pro–Gly)10 and (Pro–Hyp–Gly)10 are stable in methanol. To investigate the question of stability further, one group of investigators synthesized a synthetic collagen that contains 4-fluoroproline (Flp), which resembles hydroxyproline except that fluorine substitutes for the hydroxyl group. The melting points of three synthetic collagen molecules were measured and are shown in the table. Synthetic peptide Forms a triple helix? Tm (8C) (Pro–Pro–Gly)10 (Pro–Hyp–Gly)10 (Pro–Flp–Gly)10 Yes Yes Yes 41 60 91 (a) Compare the structure of hydroxyproline and fluoroproline. Why do you suppose the investigators chose fluorine? (b) Compare the melting points of the three synthetic collagens. What factor contributes the most to the stability of the molecule? 156 61. The triple helix of collagen is stabilized through hydrogen bonding. Draw an example of one of these hydrogen bonds, as described in the text. 62. Because collagen molecules are difficult to isolate from the connective tissue of animals, studies of collagen structure often use synthetic peptides. Would it be practical to try to purify collagen molecules from cultures of bacterial cells that have been engineered to express collagen genes? 63. Gelatin is a food product obtained from the thermal degradation of collagen. Powdered gelatin is soluble in hot water and forms a gel upon cooling to room temperature. Why is gelatin nutritionally inferior to other types of protein? 64. Papaya and pineapple contain proteolytic enzymes that can degrade collagen. (a) Cooking meat with these fruits helps to tenderize the meat. How does this work? (b) A cook decides to flavor a gelatin dessert with fresh papaya and pineapple. The fruits are added to a solution of gelatin in hot water, then the mixture is allowed to cool (see Problem 63). What is the result and how could the cook have avoided this? 65. A retrospective study conducted over 25 years examined a group of patients with similar symptoms: bruising, joint swelling, fatigue, and gum disease. Examination of patient records showed that some patients suffered from gastrointestinal diseases, poor dentition, or alcoholism, while others followed various food fads. (a) What disease afflicts this diverse group of patients? (b) Explain why the patients suffer from the symptoms listed. (c) Explain why your diagnosis is consistent with the patients’ medical histories. 66. Radioactively labeled [14C]-proline is incorporated into collagen in cultured fibroblasts. The radioactivity is detected in the collagen protein. But collagen synthesized in the presence of [14C]hydroxyproline is not radiolabeled. Explain why. 67. Lysyl hydroxylase, like prolyl hydroxylase, requires ascorbate. Draw the product of the lysyl hydroxylase reaction, a 5-hydroxylysyl residue. 68. In a collagen polypeptide consisting mostly of repeating Gly–Pro–Hyp sequences, which residue(s) is (are) most likely to be substituted by 5-hydroxylysine? 70. A small portion of the middle of the gene for the a1(II) chain of type II collagen (the entire gene contains 3421base pairs) is shown below as the sequence of the nontemplate, or coding, strand of the DNA. Also shown is the gene isolated from the skin fibroblasts of a patient with a mutant a1(II) gene. The patient was a neonate delivered at 38 weeks who died shortly after birth. The neonate had shortened limbs and a large head, with a flat nasal bridge and short bones and ribs. The symptoms are consistent with the disease hypochondrogenesis. Analysis of tissue taken from the neonate revealed a single point mutation in the a1(II) collagen gene. Normal a1(II) collagen gene: … TAACGGCGAGAAGGGAGAAGTTGGACCTCCT… Mutant a1(II) collagen gene: …TAACGGCGAGAAGGCAGAAGTTGGACCTCCT… (a) What is the sequence of the protein encoded by both the normal and mutant genes? (Note that you must first determine the correct open reading frame.) What is the change in the protein sequence? (b) How would the Tm values for normal collagen and mutant collagen compare? (c) Why did the patient die shortly after birth? 5-3 Motor Proteins 71. Is myosin a fibrous protein or a globular protein? Explain. 72. Myosin type V is a two-headed myosin that operates as a transport motor to move its attached cargo along actin filaments. Its mechanism is similar to that of muscle myosin, but it acts processively, like kinesin. The reaction cycle diagrammed here begins with both myosin V heads bound to an actin filament. ADP is bound to the leading head, and the nucleotide-binding site in the trailing head is empty. Cargo ADP Actin ATP 69. The effect of the drug minoxidil was measured in cultured human skin fibroblasts, and the data are shown in the table below. Fibroblasts were incubated with [3H]-proline or [3H]-lysine to allow the cells to incorporate the radioactively labeled amino acids into collagen chains. Following treatment with minoxidil (controls were untreated), the cells were harvested and then subjected to enzyme treatment, which hydrolyzed all cellular protein. The amount of [3H]-proline or [3H]-lysine incorporated into collagen was then measured in terms of 3H2O release. Culture conditions Prolyl hydroxylase activity (cpm, 3 H2O/mg protein) Lysyl hydroxylase activity (cpm, 3 H2O/mg protein) Control Minoxidil 13,056 13,242 12,402 1,936 (a) What is the effect of minoxidil on the cultured fibroblasts? (b) Why would this drug be effective in treating fibrosis (a skin condition associated with an accumulation of collagen)? (c) What are the potential dangers in the long-term use of minoxidil to treat hair loss? Based on your knowledge of the muscle myosin reaction cycle (Fig. 5-35), propose a reaction mechanism for myosin V, starting with the entry of ATP. How does each step of the ATP hydrolysis reaction correspond to a conformational change in myosin V? 157 73. Early cell biologists, examining living cells under the microscope, observed that the movement of certain cell constituents was rapid, linear, and targeted (that is, directed toward a particular point). (a) Why are these qualities inconsistent with diffusion as a mechanism for redistributing cell components? (b) List the minimum requirements for an intracellular transport system that is rapid, linear, and targeted. 74. Explain why the movement of myosin along an actin filament is “hopping,” whereas the movement of kinesin along a microtubule is “walking.” 76. (a) Why is rigor mortis of concern to those in the meat industry? (b) How could a forensic pathologist use the concept of rigor mortis to investigate the circumstances of an individual’s death? 77. In individuals with muscular dystrophy, a mutation in a gene for a structural protein in muscle fibers leads to progressive muscle weakening beginning in early childhood. Explain why muscular dystrophy patients typically also exhibit abnormal bone development. 78. Explain why a one-headed kinesin motor would be ineffective. 75. Rigor mortis is the stiffening of a corpse that occurs in the hours following death. Using what you know about the mechanism of motor proteins, explain what causes this stiffness.

SELECTED READINGS Brodsky, B., and Persikov, A. V., Molecular structure of the collagen triple helix, Adv. Protein Chem. 70, 310–339 (2005). [Discusses collagen structure and diseases associated with collagen mutations.] Endow, S. A., Kinesin motors as molecular machines, Bioessays 25, 1212–1219 (2003). Gardner, M. K., Hunt, A. J., Goodson, H. V., and Odde, D. J., Microtubule assembly dynamics: new insights at the nanoscale, Curr. Opin. Cell Biol. 20, 64–70 (2008). [Describes some of the techniques used to characterize microtubule formation.] O’Connell, C. B., Tyska, M. J., and Mooseker, M. S., Myosin at work: motor adaptations for a variety of cellular functions, Biochim. Biophys. Acta 1773, 615–630 (2007). Oshima, R. G., Intermediate filaments: a historical perspective, Exp. Cell Res. 313, 1981–1994 (2007). [Describes what is known about various types of intermediate filaments, including keratin.] Reisler, E., and Egelman, E. H., Actin structure and function: what we still do not understand, J. Biol. Chem. 282, 36113–36137 (2007). [Summarizes some areas of research on actin structure and its assembly into filaments.] Schechter, A. N., Hemoglobin research and the origins of molecular medicine, Blood 112, 3927–3938 (2008). [Reviews hemoglobin structure and function, including hemoglobin S.] Wilson, M. T., and Reeder, B. J., Oxygen-binding haem proteins, Exp. Physiol. 93, 128–132 (2007). [Discusses the lesser-known functions of hemoglobin, myoglobin, and some of their relatives.]

chapter 6 HOW ENZYMES WORK DOES the lock-and-key model explain enzyme action? In 1894, long before the first enzyme structure had been determined and several decades before it had been shown that enzymes are proteins, Emil Fischer proposed that a substrate molecule fits into the enzyme active site like a key in a lock. This simple idea has been widely embraced, in part because it appears to explain the exquisite substrate specificity of enzymes. Yet the lock model does not explain why some substances can fit into the keyhole but not undergo any chemical reaction. Once you learn more about how enzymes work, you will appreciate why Fischer’s lock-and-key model has been revised. [iSci/iStockphoto] THIS CHAPTER IN CONTEXT Do You Remember? Part 1 Foundations Living organisms obey the laws of thermodynamics (Section 1-3). Noncovalent forces, including hydrogen bonds, ionic interactions, and van der Waals forces, act on biological molecules (Section 2-1). Part 2 Molecular Structure and Function 6 How Enzymes Work An acid’s pK value describes its tendency to ionize (Section 2-3). The 20 amino acids differ in the chemical characteristics of their R groups (Section 4-1). Some proteins can adopt more than one stable conformation (Section 4-3). Part 3 Metabolism Part 4 Genetic Information 158 We have already seen how protein structures relate to their physiological functions. We are now ready to examine enzymes, which directly participate in the chemical reactions by which matter and energy are transformed by living cells. In this chapter, we examine the fundamental features of enzymes, including the thermodynamic underpinnings of their activity. We describe various mechanisms by which enzymes accelerate chemical reactions, focusing primarily on the digestive enzyme chymotrypsin to illustrate how different structural features influence catalytic activity. The following chapter continues the discussion of enzymes by describing how enzymatic activity is quantified and how it can be regulated. 159 What Is an Enzyme 6-1 What Is an Enzyme? In vitro, under physiological conditions, the peptide bond that links amino acids in peptides and proteins has a half-life of about 20 years. That is, after 20 years, about half of the peptide bonds in a given sample of a peptide will have been broken down through hydrolysis (cleavage by water): N R O CH C H H O CH C ⩽ H2O H R N N R CH O C O⩾ H ⩽ H ⩽ N R O CH C H Obviously, the long half-life of the peptide bond is advantageous for living organisms, since many of their structural and functional characteristics depend on the integrity of proteins. On the other hand, many proteins—some hormones, for example—must be broken down very rapidly so that their biological effects can be limited. Clearly, an organism must be able to accelerate the rate of peptide bond hydrolysis. In general, there are three ways to increase the rate of a chemical reaction, including hydrolysis: 1. Increasing the temperature (adding energy in the form of heat). Unfortunately, this is not very practical, since the vast majority of organisms cannot regulate their internal temperature and thrive only within relatively narrow temperature ranges. Furthermore, an increase in temperature accelerates all chemical reactions, not just the desired reaction. 2. Increasing the concentrations of the reacting substances. Higher concentrations of reactants increase the likelihood that they will encounter each other in order to react. But a cell may contain tens of thousands of different types of molecules, space is limited, and many essential reactants are scarce inside as well as outside the cell. 3. Adding a catalyst. A catalyst is a substance that participates in the reaction yet emerges at the end in its original form. A huge variety of chemical catalysts are known. For example, the catalytic converter in an automobile engine contains a mixture of platinum and palladium that accelerates the conversion of carbon monoxide and unburned hydrocarbons to the relatively harmless carbon dioxide. Living systems use catalysts called enzymes to increase the rates of chemical reactions. Most enzymes are proteins, but a few are made of RNA (these are called ribozymes and are described more fully in Section 21-3). One of the best-studied enzymes is chymotrypsin, a digestive protein that is synthesized in the pancreas and secreted KEY CONCEPTS • Enzymes differ from simple chemical catalysts in their efficiency and specificity. • An enzyme’s name may reflect the reaction it catalyzes. Figure 6-1 Ribbon model of chymotrypsin. The polypeptide chain (gray) folds into two domains. Three residues essential for the enzyme’s activity are shown in red. [Structure (pdb 4CHA) determined by H. Tsukada and D. M. Blow.] into the small intestine, where it helps break down dietary proteins. Perhaps because it can be purified in relatively large quantities from the pancreas of cows, chymotrypsin was one of the first enzymes to be crystallized (it is also widely used in the laboratory; for example, see Table 4-4). Chymotrypsin’s 241 amino acid residues form a compact two-domain structure (Fig. 6-1). Hydrolysis of polypeptide substrates takes place in a cleft between the two domains, near the side chains of three residues (His 57, Asp 102, and Ser 195). This area of the enzyme is known as the active site. The active sites of nearly all known enzymes are located in similar crevices on the enzyme surface. Chymotrypsin catalyzes the hydrolysis of peptide bonds at a rate of about 190 per second, which is about 1.7 3 1011 times faster than in the absence of a catalyst. This is also orders of magnitude faster than would be possible with a simple chemical catalyst. In addition, chymotrypsin and other enzymes act under mild conditions (atmospheric pressure and physiological temperature), whereas many chemical catalysts require extremely high temperatures and pressures for optimal performance. Chymotrypsin’s catalytic power is not unusual: Rate enhancements of 10 8 to 10 12 are typical of enzymes (Table 6-1 gives the rates of some enzyme-catalyzed reactions). Of course, the slower the rate of the uncatalyzed reaction, the greater the opportunity for rate enhancement by an enzyme (see, for example, orotidine-59-monophosphate decarboxylase in Table 6-1). Interestingly, even relatively fast reactions are subject to enzymatic catalysis in biological systems. For example, the conversion of CO2 to carbonic acid in water CO2 1 H2O Δ H2CO3 has a half-time of 5 seconds (half the molecules will have reacted within 5 seconds). This reaction is accelerated over a millionfold by the enzyme carbonic anhydrase (see Table 6-1). Another feature that sets enzymes apart from nonbiological catalysts is their reaction specificity. Most enzymes are highly specific for their reactants (called substrates) and products. The functional groups in the active site of an enzyme are so carefully arranged that the enzyme can distinguish its substrates from among many others that are similar in size and shape and can then mediate a single chemical reaction involving those substrates. This reaction specificity stands in marked contrast to the permissiveness of most organic catalysts, which can act on many different kinds of substrates and, for a given substrate, sometimes yield more than one product. Chymotrypsin and some other digestive enzymes are somewhat unusual in acting on a relatively broad range of substrates and, at least in the laboratory, catalyzing several types of reactions. For instance, chymotrypsin catalyzes the hydrolysis of the peptide bond following almost any large nonpolar residue such as Phe, Trp, or Tyr. It can also catalyze the hydrolysis of other amide bonds and ester bonds. This behavior TABLE 6-1 Rate Enhancements of Enzymes Enzyme Orotidine-59monophosphate decarboxylase Staphylococcal nuclease Adenosine deaminase Chymotrypsin Triose phosphate isomerase Chorismate mutase Carbonic anhydrase a Uncatalyzed Rate (s21) 78,000,000 years 2.8 3 10216 39 1.4 3 1017 1.7 3 10213 1.8 3 10210 1.0 3 1029 4.3 3 1026 2.6 3 1025 1.3 3 1021 95 370 190 4,300 50 1,000,000 5.6 3 1014 2.1 3 1012 1.7 3 1011 1.0 3 109 1.9 3 106 7.7 3 106 130,000 120 20 1.9 7.4 5 years years years years hours seconds The half-times of very slow reactions were estimated by extrapolating from measurements made at very high temperatures. [Data mostly from Radzicka, R., and Wolfenden, R., Science 267, 90–93 (1995).] Catalyzed Rate (s21) Rate Enhancement (catalyzed rate/ uncatalyzed rate) Half-Time (uncatalyzed)a has proved to be convenient for quantifying the activity of purified chymotrypsin. An artificial substrate such as p-nitrophenylacetate (an ester) is readily hydrolyzed by the action of chymotrypsin (the name of the enzyme appears next to the reaction arrow to indicate that it participates as a catalyst): 161 What Is an Enzyme O H3C C O NO2 p-Nitrophenylacetate (colorless) H2O chymotrypsin 2 H⩽ O H3C C O⩾ ⩽ ⩾ O NO2 p-Nitrophenolate (yellow) Acetate The p-nitrophenolate reaction product is bright yellow, so the progress of the reaction can be easily monitored in a spectrophotometer at 410 nm. Finally, enzymes differ from nonbiological catalysts in that the activities of many enzymes are regulated so that the organism can respond to changing conditions or follow genetically determined developmental programs. For this reason, biochemists seek to understand how enzymes work as well as when and why. These aspects of enzyme behavior are fairly well understood for chymotrypsin, which makes it an ideal subject to showcase the fundamentals of enzyme activity. Enzymes are usually named after the reaction they catalyze The enzymes that catalyze biochemical reactions have been formally classified into six subgroups according to the type of reaction carried out (Table 6-2). Basically, all biochemical reactions involve either the addition of some substance to another, or its removal, or the rearrangement of that substance. Keep in mind that although the substrates of many biochemical reactions appear to be quite large (for example, proteins or nucleic acids), the action really involves just a few chemical bonds and a few small groups (sometimes H2O or even just an electron). The name of an enzyme frequently provides a clue to its function. In some cases, an enzyme is named by incorporating the suffix -ase into the name of its substrate. For example, fumarase is an enzyme that acts on fumarate (Reaction 7 in the citric acid cycle; see Section 14-2). Chymotrypsin can similarly be called a proteinase, a protease, or a peptidase. Most enzyme names contain more

descriptive words (also ending in -ase) to indicate the nature of the reaction catalyzed by that enzyme. For example, pyruvate decarboxylase catalyzes the removal of a CO2 group from pyruvate: O C C H⩽ O CH3 Pyruvate Enzyme Classification Class of Enzyme 1. Oxidoreductases 2. Transferases 3. Hydrolases 4. Lyases ⩾ O TABLE 6-2 pyruvate decarboxylase H O C CH3 Acetaldehyde 5. Isomerases ⩽ CO2 6. Ligases Type of Reaction Catalyzed Oxidation–reduction reactions Transfer of functional groups Hydrolysis reactions Group elimination to form double bonds Isomerization reactions Bond formation coupled with ATP hydrolysis 162 Ch 6 How Enzymes Work Alanine aminotransferase catalyzes the transfer of an amino group from alanine to an a-keto acid: ⩽ O NH3 H 3C COO⩾ ⩽ CH ⩾ OOC CH2 C CH2 COO⩾ ␣-Ketoglutarate Alanine alanine aminotransferase ⩽ O H3C C NH3 COO⩾ ⩽ ⩾ OOC CH Pyruvate CH2 CH2 COO⩾ Glutamate Such a descriptive naming system tends to break down in the face of the many thousands of known enzyme-catalyzed reactions, but it is adequate for the small number of well-known reactions that are included in this book. A more precise classification scheme systematically groups enzymes in a four-level hierarchy and assigns each enzyme a unique number. For example, chymotrypsin is known as EC 3.4.21.1 (EC stands for Enzyme Commission, part of the nomenclature committee of the International Union of Biochemistry and Molecular Biology; the EC database can be accessed at). Keep in mind that even within an organism, more than one protein may catalyze a given chemical reaction. Multiple enzymes catalyzing the same reaction are called isozymes. Although they usually share a common evolutionary origin, isozymes differ in their catalytic properties. Consequently, the various isozymes that are expressed in different tissues or at different developmental stages can perform slightly different metabolic functions. CONCEPT REVIEW • How do enzymes differ from nonbiological catalysts? • Explain why an enzyme's common name may not reveal its biological role. 6-2 The Chemistry of Catalysis KEY CONCEPTS • The height of the activation energy barrier determines the rate of a reaction. • An enzyme provides a lower-energy pathway from reactants to products. • Enzymes accelerate chemical reactions using acid–base catalysis, covalent catalysis, and metal ion catalysis. • Chymotrypsin's catalytic triad participates in acid–base and covalent catalysis. In a biochemical reaction, the reacting species must come together and undergo electronic rearrangements that result in the formation of products. In other words, some old bonds break and new bonds form. Let us consider an idealized transfer reaction in which compound AOB reacts with compound C to form two new compounds, A and BOC: A B + C → A + B C In order for the first two compounds to react, they must approach closely enough for their constituent atoms to interact. Normally, atoms that approach too closely repel each other. But if the groups have sufficient free energy, they can pass this point and react with each other to form products. The progress of the reaction can be depicted on a diagram (Fig. 6-2) in which the horizontal axis represents the progress of the reaction (the reaction coordinate) and the vertical axis represents the free energy (G) of the system. The energy-requiring step of the reaction is shown as an energy barrier, called the free energy of activation or activation energy and symbolized DG ‡. The point of highest energy is known as the transition state and can be considered a sort of intermediate between the reactants and products. The lifetimes of transition states are extremely short, on the order of 10214 to 213 seconds. Because they are too short-lived to be accessible to most analytical 10 163 The Chemistry of Catalysis Progress of reaction A B Free energy (G) C Figure 6-2 Reaction coordinate diagram for the reaction AOB 1 C n A 1 BOC. The progress of the reaction + ΔG + A B+C A+B is shown on the horizontal axis, and free energy is shown on the vertical axis. The transition state of the reaction, represented as A…B…C, is the point of highest free energy. The free energy difference between the reactants and the transition state is the free energy of activation (DG ‡). C Reaction coordinate techniques, the transition states of many reactions cannot be identified with absolute certainty. However, it is useful to visualize the transition state as a molecular species in the process of breaking old bonds and forming new bonds. For the reaction above, we can represent this as A????B????C. The reactants require free energy (DG ‡) to reach this point (for example, some energy is required to break existing bonds and bring other atoms together to begin forming new bonds). The analogy of going uphill in order to undergo a reaction is appropriate. The height of the activation energy barrier determines the rate of a reaction (the amount of product formed per unit time). The higher the activation energy barrier, the less likely the reaction is to occur (the slower it is). Although the reactant molecules have varying free energies, very few of them have enough free energy to reach the transition state during a given time interval. But the lower the energy barrier, the more likely the reaction is to occur (the faster it is), because more reactant molecules happen to have enough free energy to achieve the transition state during the same time interval. Note that the transition state, at the peak, can potentially roll down either side of the free energy hill. Therefore, not all the reactants that get together to form a transition state actually proceed all the way to products; they may return to their original state. Similarly, the products (A and B—C in this case) can react, pass through the same transition state (A????B????C), and yield the original reactants (AOB and C). In nature, the free energies of the reactants and products of a chemical reaction are seldom identical, so the reaction coordinate diagram looks more like Figure 6-3a. When the products have a lower free energy than the reactants, then the overall free (a) (b) Free energy (G) Reactants ΔG reaction Products Reaction coordinate Free energy (G) Products ΔG reaction Reactants Reaction coordinate Figure 6-3 Reaction coordinate diagram for a reaction in which reactants and products have different free energies. The free energy change for the reaction (DG) is equivalent to Gproducts 2 Greactants. (a) When the free energy of the reactants is greater than that of the products, the free energy change for the reaction is negative, so the reaction proceeds spontaneously. (b) When the free energy of the products is greater than that of the reactants, the free energy change for the reaction is positive, so the reaction does not proceed spontaneously (however, the reverse reaction does proceed). ? In a cell, some enzyme-catalyzed reactions proceed in both the forward and reverse directions. Sketch a reaction coordinate diagram for such a reaction. 164 Ch 6 How Enzymes Work energy change of the reaction (DGreaction, or Gproducts 2 Greactants) is less than zero. A negative free energy change indicates that a reaction proceeds spontaneously as written. Note that “spontaneously” does not mean “quickly.” A reaction with a negative free energy change is thermodynamically favorable, but the height of the activation energy barrier (DG ‡) determines how fast the reaction actually occurs. If the products have greater free energy than the reactants (Fig. 6-3b), then the overall free energy change for the reaction (DGreaction) is greater than zero. This reaction does not proceed as written (because it goes “uphill”), but it does proceed in the reverse direction. A catalyst provides a reaction pathway with a lower activation energy barrier A catalyst, whether inorganic or enzymatic, decreases the activation energy barrier (DG ‡) of a reaction (Fig. 6-4). It does so by interacting with the reacting molecules such that they are more likely to assume the transition state. A catalyst speeds up the reaction because as more reacting molecules achieve the transition state per unit time, more molecules of product can form per unit time. (An increase in temperature increases the rate of a reaction for a similar reason: The input of thermal energy allows more molecules to achieve the transition state per unit time.) Thermodynamic calculations indicate that lowering DG ‡ by about 5.7 kJ ? mol21 accelerates the reaction 10-fold. A rate increase of 106 requires lowering DG ‡ by six times this amount, or about 34 kJ ? mol21. An enzyme catalyst does not alter the net free energy change for a reaction; it merely provides a pathway from reactants to products that passes through a transition state that has lower free energy than the transition state of the uncatalyzed reaction. An enzyme, therefore, lowers the height of the activation energy barrier (DG ‡) by lowering the energy of the transition state. The hydrolysis of a peptide bond is always thermodynamically favorable, but the reaction occurs quickly only when a catalyst (such as the protease chymotrypsin) is available to provide a lower-energy route to the transition state. Enzymes use chemical catalytic mechanisms The idea that living organisms contain agents that can promote the change of one substance into another has been around at least since the early nineteenth century, when scientists began to analyze the chemical transformations carried out by organisms such as yeast. However, it took some time to appreciate that these catalytic agents were not part of some “vital force” present only in intact, living organisms. In 1878, the word enzyme was coined to indicate that there was something in yeast Figure 6-4 Effect of a catalyst + X +uncat on a chemical reaction. Here, the reactants proceed through a transition state symbolized X‡ during their conversion to products. A catalyst lowers the free energy of activation ‡ (DG ‡) for the reaction so that DG cat , ‡ ,DG uncat. Lowering the free energy of the transition state (X‡) accelerates the reaction because more reactants are able to achieve the transition state per unit time. Uncatalyzed reaction + ΔG +uncat ++ X cat Free energy (G) Catalyzed reaction + + ΔG cat Reactants Products Reaction coordinate (Greek en 5 “in,” zyme 5 “yeast”), rather than the yeast itself, that was responsible for breaking down (fermenting) sugar. In fact, the action of enzymes can be explained in purely chemical terms. What we currently know about enzyme mechanisms Metal ions rests solidly on a foundation of knowledge about simple chemical catalysts. In an enzyme, certain functional groups in the enzyme's active site perform the same catalytic function as small chemical catalysts. In some cases, the amino acid side chains of an enzyme cannot provide the required catalytic groups, so a tightly bound cofactor participates in catalysis. For example, many oxidation–reduction reactions require a metal ion cofactor, since a metal ion can exist in multiple oxidation states, unlike an amino acid side chain. Some enzyme cofactors are organic molecules known as coenzymes, which may be derived from vitamins. Enzymatic activity still requires the protein portion of the enzyme, which helps position the cofactor and reactants for the reaction (this situation is reminiscent of myoglobin and hemoglobin, where the globin and heme group together function to bind oxygen; see Section 5-1). Some coenzymes, termed cosubstrates, enter and exit the active site as substrates do; a tightly bound coenzyme that remains in the active site between reactions is called a prosthetic group (Fig. 6-5). There are three basic kinds of chemical catalytic mechanisms used by enzymes: acid–base catalysis, covalent catalysis, and metal ion catalysis. We will examine each of these, using model reactions to illustrate some of their fundamental features. 1. Acid–Base Catalysis Many enzyme mechanisms include acid–base catalysis, in which a proton is transferred between the enzyme and the substrate. This mechanism of catalysis can be further divided into acid catalysis and base catalysis. Some enzymes use one or the other; many use both. Consider the following model reaction, the tautomerization of a ketone to an enol (tautomers are interconvertible isomers that differ in the placement of a hydrogen and a double bond): R R C O R O⩾ C CH2 CH2 H H⩽ Ketone C O H CH2 Transition state Enol Here the transition state is shown in square brackets to indicate that it is an unstable, transient species. The dotted lines represent bonds in the process of breaking or forming. The uncatalyzed reaction occurs slowly because formation of the carbanion-like transition state has a high activation energy barrier (a carbanion is a compound in which the carbon atom bears a negative charge). If a catalyst (symbolized HOA) donates a proton to the ketone's oxygen atom, it reduces the unfavorable carbanion character of the transition state, thereby lowering its energy and hence lowering the activation energy barrier for the reaction: R R C O ⩽ H A C R O⩾ CH2 CH2 H H⩽ H⩽ A⩾ C O H ⩽ H A CH2 This is an example of acid catalysis, since the catalyst acts as an acid by donating a proton. Note that the catalyst is returned to its original form at the end of the reaction. 165 Cofactors Coenzymes Cosubstrates Prosthetic groups Figure 6-5 Types of enzyme cofactors. 166 Ch 6 How Enzymes Work O Asp CH2 C Glu CH2 CH2 His CH2 OH O Figure 6-6 Amino acid side chains that can act as acid–base catalysts. These groups can act as acid or base catalysts, depending on their state of protonation in the enzyme's active site. The side chains are shown in their protonated forms, with the acidic proton highlighted. ? Lys CH2 C OH ⩽ NH N H CH2 H CH2 CH2 ⩽ N H H At neutral pH, which of these side chains most likely function as acid catalysts? Which most likely function as base catalysts? (Hint: See Table 4-1.) Cys CH2 Tyr CH2 SH OH The same keto-enol tautomerization reaction shown above can be accelerated by a catalyst that can accept a proton, that is, by a base catalyst. Here, the catalyst is shown as :B, where the dots represent unpaired electrons: R H⩽ R O ⩽ B C C CH2 CH2 H H⩽ R O⩾ C O H ⩽ B ⩽ H⩽ CH2 B Base catalysis lowers the energy of the transition state and thereby accelerates the reaction. In enzyme active sites, several amino acid side chains can potentially act as acid or base catalysts. These are the groups whose pK values are in or near the physiological pH range. The residues most commonly identified as acid–base catalysts are shown in Figure 6-6. Because the catalytic functions of these residues depend on their state of protonation or deprotonation, the catalytic activity of the enzyme may be sensitive to changes in pH. 2. Covalent Catalysis In covalent catalysis, the second major chemical reaction mechanism used by enzymes, a covalent bond forms between the catalyst and the substrate during formation of the transition state. Consider as a model reaction the decarboxylation of acetoacetate. In this reaction, the movement of electron pairs among atoms is indicated by red curved arrows (Box 6-A). O H3C C O CH2 Acetoacetate C O⩾ ⩾ CO2 H3C O C Enolate O H⩽ CH2 H3C C Acetone CH3 167 The Chemistry of Catalysis BIOCHEMISTRY NOTE BOX 6-A Depicting Reaction Mechanisms While it is often sufficient to draw the structures of a reaction's substrates and products, a full understanding of the reaction mechanism requires knowing what the electrons are doing. Recall that a covalent bond forms when two atoms share a pair of electrons, and a vast number of biochemical reactions involve breaking and forming covalent bonds. Although single-electron reactions also occur in biochemistry, we will focus on the more common two-electron reactions. The curved arrow convention shows how electrons are rearranged during a reaction. The arrow emanates from the original location of an electron pair. This can be either an unshared electron pair, on an atom such as N or O, or the electrons of a covalent bond. The curved arrow points to the final location of the electron pair. For example, to show a bond breaking: X X⩽ ⩽ Y Y⩾ and to show a bond forming: X⩽ ⩽ SY ⩾ X Y A familiarity with Lewis dot structures and an understanding of electronegativity (see Section 2-1) is helpful for identifying electron-rich groups (the source of electrons during a reaction) and electron-poor groups (where the electrons often end up). A typical biochemical reaction requires several curved arrows, for example, R O H C O O SN NH ⩽ NH H R R ⩾ O C OH ⩽ H N O R Question: Draw arrows to show the electron movements in the reaction 2 H2O n OH2 1 H3O1. The transition state, an enolate, has a high free energy of activation. This reaction can be catalyzed by a primary amine (RNH2), which reacts with the carbonyl group of acetoacetate to form an imine, a compound containing a CPN bond (this adduct is known as a Schiff base): R O H3C C O CH2 Acetoacetate C O⩾ ⩽ RNH2 H3C ⩽ H C CH2 N O C Schiff base (imine) O⩾ ⩽ OH⩾ 168 Ch 6 How Enzymes Work ++ X2 ++ X1 ++ ΔG 2 Intermediate Free energy (G) ++ ΔG 1 Reactants Products Reaction coordinate Figure 6-7 Diagram for a reaction accelerated by covalent catalysis. Two transition states flank the covalent intermediate. The relative heights of the activation energy barriers (to achieve the two transition states, X‡1 and X‡2) vary depending on the reaction. ? Explain why the free energy of the reaction intermediate must be greater than the free energy of either the reactants or products. In this covalent intermediate, the protonated nitrogen atom acts as an electron sink to reduce the enolate character of the transition state in the decarboxylation reaction: R H3C H ⩽ N R O C CH2 C CO2 H3C O⩾ N C H ⩽ R H CH2 H 3C ⩽ H C CH3 N Finally, the Schiff base decomposes, which regenerates the amine catalyst and releases the product, acetone: R H3C ⩽ N C H CH3 ⩽ OH Schiff base Ser, Tyr OH O Cys SH S Lys NH3 ⩽ ⩽ N H ⩾ NH2 N NH His ⩾ N H Figure 6-8 Protein groups that can act as covalent catalysts. In their deprotonated forms (right), these groups act as nucleophiles. They attack electron-deficient centers to form covalent intermediates. O ⩾ H3C C CH3 ⩽ RNH2 Acetone In enzymes that use covalent catalysis, an electron-rich group in the enzyme active site forms a covalent adduct with a substrate. This covalent complex can sometimes be isolated; it is much more stable than a transition state. Enzymes that use covalent catalysis undergo a two-part reaction process, so the reaction coordinate diagram contains two energy barriers with the reaction intermediate between them (Fig. 6-7). Many of the same groups that make good acid–base catalysts (see Fig. 6-6) also make good covalent catalysts because they contain unshared electron pairs (Fig. 6-8). Covalent catalysis is often called nucleophilic catalysis because the catalyst is a nucleophile, that is, an electron-rich group in search of an electron-poor center (a compound with an electron deficiency is known as an electrophile). 3. Metal Ion Catalysis Metal ion catalysis occurs when metal ions participate in enzymatic reactions by mediating oxidation–reduction reactions, as mentioned earlier, or by promoting the reactivity of other groups in the enzyme's active site through electrostatic effects. A protein-bound metal ion can also interact directly with the reacting substrate. For example, during the conversion of acetaldehyde to ethanol as catalyzed by the liver enzyme alcohol dehydrogenase, a zinc ion stabilizes the developing negative charge on the oxygen atom during formation of the transition state: H3C C O H3C H C O⩾ Zn2⩽ OH 2H H3C H 169 The Chemistry of Catalysis C H H Acetaldehyde Ethanol The catalytic triad of chymotrypsin promotes peptide bond hydrolysis Chymotrypsin uses both acid–base catalysis and covalent catalysis to accelerate peptide bond hydrolysis. These activities depend on three active-site residues whose identities and catalytic importance have been the object of intense study since the 1960s. Two of chymotrypsin's catalytic residues were identified using a technique called chemical labeling. When chymotrypsin is incubated with the compound diisopropylphosphofluoridate (DIPF), one of its 27 Ser residues (Ser 195) becomes covalently tagged with the diisopropylphospho (DIP) group, and the enzyme loses activity. CH(CH3)2 HF O CH2 OH ⩽ F P O O CH(CH3)2 Ser 195 (active) CH(CH3)2 Diisopropylphosphofluoridate (DIPF) O CH2 O P O O CH(CH3)2 DIP-Ser 195 (inactive) This observation provided strong evidence that Ser 195 is essential for catalysis. Chymotrypsin is therefore known as a serine protease. It is one of a large family of enzymes that use the same Ser-dependent catalytic mechanism. A similar labeling technique was used to identify the catalytic importance of His 57. The third residue involved in catalysis by chymotrypsin—Asp 102—was identified only after the fine structure of chymotrypsin was visualized through X-ray crystallography. The hydrogen-bonded arrangement of the Asp, His, and Ser residues of chymotrypsin and other serine proteases is called the catalytic triad (Fig. 6-9). The substrate's scissile bond (the bond to be cleaved by hydrolysis) is positioned near Ser 195 when the substrate binds to the enzyme. The side chain of Ser is not normally a strong enough nucleophile to attack an amide bond. However, His 57, acting as a base catalyst, abstracts a proton from Ser 195 so that the oxygen can act as a covalent catalyst. Asp 102 promotes catalysis by stabilizing the resulting positively charged imidazole group of His 57. Chymotrypsin-catalyzed peptide bond hydrolysis actually occurs in two phases that correspond to the formation and breakHis 57 down of a covalent reaction intermediate. The steps of catalysis Asp 102 are detailed in Figure 6-10. Nucleophilic attack of Ser 195 on the substrate's carbonyl carbon leads to an unstable tetrahedral intermediate in which the carbonyl carbon assumes tetrahedral geometry. This structure then collapses to an intermediate in which the Ser 195 N-terminal portion of the substrate remains covalently attached Figure 6-9 The catalytic triad of chymotrypsin. Asp to the enzyme. The second part of the reaction, during which the 102, His 57, and Ser 195 are arrayed in a hydrogen-bonded oxygen of a water molecule attacks the carbonyl carbon, also in- network. Atoms are color-coded (C gray, N blue, O red), cludes an unstable tetrahedral intermediate. Although the enzyme- and the hydrogen bonds are shaded yellow. catalyzed reaction requires multiple steps, the net reaction is the same as the uncatalyzed reaction shown in Section 6-1. ? Add hydrogen atoms to the three side chains. His 57 Ser 195 CH2 O ⩾ C O H N CH2 CH2 N H O RC N C Asp 102 H O RN The peptide substrate enters the active site of chymotrypsin so that its scissile bond (red) is close to the oxygen of Ser 195 (the N-terminal portion of the substrate is represented by RN, and the C-terminal portion by R C). 1. Removal of the Ser hydroxyl proton by His 57 (a base catalyst) allows the resulting nucleophilic oxygen (a covalent catalyst) to attack the carbonyl carbon of the substrate. His 57 ⩾ C O H N CH2 Asp 102 N H ⩽ CH2 Asp 102 C O H N CH2 O H N N CH2 N C O RN Ser 195 CH2 O ⩾ C O H N CH2 ⩽ N H O O C CH2 Asp 102 H O⩾ H RN Tetrahedral intermediate O Ser 195 C 5. The N-terminal portion of the original substrate, now with a new C-terminus, diffuses away, regenerating the enzyme. O RN CH2 O C O His 57 2. The tetrahedral intermediate decomposes when His 57, now acting as an acid catalyst, donates a proton to the nitrogen of the scissile peptide bond. This step cleaves the peptide bond. Asp 102 promotes the reaction by stabilizing His 57 through hydrogen bonding. N O 4. In the second tetrahedral intermediate, His 57, now an acid catalyst, donates a proton to the Ser oxygen, leading to collapse of the intermediate. This step resembles Step 2 above. O⩾ C H H N H O CH2 C ⩾ CH2 His 57 ⩾ O H RN Tetrahedral intermediate RC Ser 195 CH2 Asp 102 RC H O His 57 Ser 195 CH2 O 3. Water then enters the active site. It donates a proton to His 57 (again a base catalyst), leaving a hydroxyl group that attacks the carbonyl group of the remaining substrate. This step resembles Step 1 above. H 2O His 57 O Ser 195 CH2 O RN Acyl–enzyme intermediate (covalent intermediate) ⩾ C The departure of the C-terminal portion of the cleaved peptide, with a newly exposed N-terminus, leaves the N-terminal portion of the substrate (an acyl group) linked to the enzyme. This relatively stable covalent complex is known as the acyl–enzyme intermediate. O H N CH2 N H O CH2 Asp 102 Figure 6-10 The catalytic mechanism of chymotrypsin and other serine proteases. The two tetrahedral intermediates are believed to resemble the transition states leading to and from the acyl–enzyme intermediate. ? 170 Choose one step of the reaction and identify all the bonds that break and form in that step. The roles of Asp, His, and Ser in peptide bond hydrolysis, as catalyzed by chymotrypsin and other members of the serine protease family, have been tested through site-directed mutagenesis (see Section 3-4). Replacing the catalytic Asp with another residue decreases the rate of substrate hydrolysis about 5000-fold. Adding a methyl group to His by chemical labeling (so that it can't accept or donate a proton) has a similar effect. Replacing the catalytic Ser with another residue decreases enzyme activity about a millionfold. Surprisingly, replacing all three catalytic residues—Asp, His, and Ser—through site-directed mutagenesis does not completely abolish protease activity: The modified enzyme still catalyzes peptide bond hydrolysis at a rate about 50,000 times greater than the rate of the uncatalyzed reaction. Clearly, chymotrypsin and its relatives rely on the acid–base catalysis and covalent catalysis carried out by the Asp–His–Ser catalytic triad, but these enzymes must have additional catalytic mechanisms that allow them to achieve reaction rates 1011 times greater than the rate of the uncatalyzed reaction. 171 The Unique Properties of Enzyme Catalysts CONCEPT REVIEW • Why does a free energy barrier separate reactants and products in a chemical reaction? • What is the relationship between the height of the activation energy barrier and the rate of the reaction? • How does an enzyme lower the activation energy of a reaction? • How can an acid or base accelerate a reaction? • How does a covalent catalyst accelerate a reaction? • Why must a reaction that proceeds via covalent catalysis have two transition states? • How do metal ions accelerate reactions? • How can chemical labeling identify an enzyme's catalytic residues? • Which residues constitute chymotrypsin's catalytic triad? • How do these residues participate in peptide bond hydrolysis as catalyzed by chymotrypsin? 6-3 The Unique Properties of Enzyme Catalysts If only a few residues in an enzyme directly participate in catalysis (for example, Asp, His, and Ser in chymotrypsin), why are enzymes so large? One obvious answer is that the catalytic residues must be precisely aligned in the active site, so a certain amount of surrounding structure is required to hold them in place. This view supports Fischer's lock-and-key model for enzyme action. But how can an active site that perfectly accommodates substrates also accommodate products before they are released from the enzyme? Moreover, how can the lock-and-key model explain why an enzyme inhibitor can bind tightly in the active site but not react? The answer is that enzymes, like other proteins (see Section 4-3), are not rigid molecules but instead can flex while binding substrates. In other words, the enzyme–substrate interaction must be more dynamic than a key in a lock, more like a hand in a glove. In some cases, the enzyme can physically distort the substrate as it binds, in effect pushing it toward a higher-energy conformation closer to the reaction's transition state. Current theories attribute much of the catalytic power of enzymes to these and other specific interactions between enzymes and their substrates. Enzymes stabilize the transition state The lock-and-key model does contain a grain of truth, a principle first formulated by Linus Pauling in 1946. He proposed that an enzyme increases the reaction rate not by binding tightly to the substrates but by binding tightly to the reaction's transition state (that is, substrates that have been strained toward the structures of the products). In other words, the tightly bound key of the lock-and-key model is the transition state, KEY CONCEPT • The catalytic activity of enzymes also depends on transition state stabilization, proximity and orientation effects, induced fit, and electrostatic catalysis. See Guided Exploration. Catalytic mechanism of serine proteases. DOES the lock-and-key model explain enzyme action? 172 Ch 6 How Enzymes Work not the substrate. In an enzyme, tight binding (stabilization) of the transition state occurs in addition to acid–base, covalent, or metal ion catalysis. In general, transition state stabilization is accomplished through the close complementarity in shape and charge between the active site and the transition state. A nonreactive substance that mimics the transition state can therefore bind tightly to the enzyme and block its catalytic activity (enzyme inhibition is discussed further in Section 7-3). Transition state stabilization appears to be an important part of the chymotrypsin reaction. In this case, the two tetrahedral intermediates, which resemble the transition state (see Fig. 6-10), are stabilized through interactions that do not occur at any other point in the reaction. Rate acceleration is believed to result from an increase in both the number and strength of the bonds that form between active site groups and the substrate in the transition state. 1. During formation of the tetrahedral intermediate, the planar peptide group of the substrate changes its geometry, and the carbonyl oxygen, now an anion, moves into a previously unoccupied cavity near the Ser 195 side chain. In this cavity, called the oxyanion hole, the substrate oxygen ion can form two new hydrogen bonds with the backbone NH groups of Ser 195 and Gly 193 (Fig. 6-11). The backbone NH group of the substrate residue preceding the scissile bond forms another hydrogen bond to Gly 193 (not shown in Fig. 6-11). Thus, the transition state is stabilized (its energy lowered) by three hydrogen bonds that cannot form when the enzyme first binds its substrate. The stabilizing effect of these three new hydrogen bonds could account for a significant portion of chymotrypsin's catalytic power, since the energy of a standard hydrogen bond is about 20 kJ ? mol21 and the reaction rate increases 10-fold for every decrease in DG ‡ of 5.7 kJ ? mol21. His 57 (a) Ser 195 C␣ CH2 CH2 O ⩾ C Asp 102 O H N N H O R O N C H R CH2 C␣ CH2 CH2 O ⩾ C Asp 102 O H N CH2 H H N Gly 193 Ser 195 His 57 (b) N ⩽ N H O R N C H R O⩾ N H H N Gly 193 Figure 6-11 Transition state stabilization in the oxyanion hole. (a) The chymotrypsin active site is shown with the oxyanion hole shaded in pink. The carbonyl carbon of the peptide substrate has trigonal geometry, so the carbonyl oxygen cannot occupy the oxyanion hole. (b) Nucleophilic attack by the oxygen of Ser 195 on the substrate carbonyl group leads to a transition state, in which the carbonyl carbon assumes tetrahedral geometry. At this point, the substrate's anionic oxygen (the oxyanion) can move into the oxyanion hole, where it forms hydrogen bonds (shaded yellow) with two enzyme backbone groups. See Animated Figure. Effect of preferential transition state binding. 2. NMR studies, which can identify individual hydrogen-bonding interactions, suggest that the hydrogen bond between Asp 102 and His 57 becomes shorter during formation of the two tetrahedral intermediates (Fig. 6-12). Such a bond is called a low-barrier hydrogen bond because the hydrogen is shared equally between the original donor and acceptor atoms (in a standard hydrogen bond, the proton still “belongs” to the donor atom and there is an energy barrier for its full transfer to the acceptor atom). A decrease in bond length from ,2.8 Å to ,2.5 Å in forming the low-barrier hydrogen bond is accompanied by a three- to fourfold increase in bond strength. The low-barrier hydrogen bond that forms during catalysis in chymotrypsin helps stabilize the transition state and thereby accelerate Asp 102 the reaction. Efficient catalysis depends on proximity and orientation effects 173 The Unique Properties of Enzyme Catalysts His 57 CH2 O C ⩾ O H N N CH2 His 57 CH2 Enzymes increase reaction rates by bringing reacting groups into close proximity O ⩾ so as to increase the frequency of collisions that can lead to a reaction. FurtherC O H N ⩽ N H more, when substrates bind to an enzyme, their translational and rotational motions are frozen out so that they can be oriented properly for reaction Asp 102 CH2 (Fig. 6-13). These proximity and orientation effects likely explain some of the Figure 6-12 Formation of a low-barrier residual activity of chymotrypsin whose catalytic residues have been altered. hydrogen bond during catalysis in Nevertheless, an enzyme must be more than a template for assembling and aligning chymotrypsin. The Asp 102—His 57 reacting groups. hydrogen bond becomes shorter and stronger, so the imidazole proton comes to be shared equally between the O of The active-site microenvironment promotes catalysis Asp and the N of histidine in a lowbarrier hydrogen bond. In nearly all cases, an enzyme's active site is somewhat removed from the solvent, with its catalytic residues in some sort of cleft or pocket on the enzyme surface. Upon binding substrates, some enzymes undergo a pronounced conformational change so that they almost fully enclose the substrates. Daniel Koshland has called this phenomenon induced fit. Many studies of enzyme activity, supplemented by X-ray crystallographic data, support this feature of enzyme action, further highlighting the shortcomings of the lock-and-key model. In fact, techniques that assess protein structure on millisecond time scales suggest that conformational fluctuations may be critical for all stages of an enzyme-catalyzed reaction, not just substrate binding. Reactants Enzyme (a) (b) Figure 6-13 Proximity and orientation effects in catalysis. In order to react, two groups must come together and collide with the correct orientation. (a) Reactants in solution are separated in space and have translational and rotational motions that must be overcome. (b) When the reactants bind to an enzyme, their motion is limited, and they are held in close proximity and with the correct alignment for a productive reaction. 174 Ch 6 How Enzymes Work A classic case of induced fit occurs in hexokinase, which catalyzes the phosphorylation of glucose by ATP (Reaction 1 of glycolysis; Section 13-1): HO O H HO ⩾2O CH2 OH H H OH 3P H H ⩽ ATP OH hexokinase Glucose (a) O HO CH2 O H OH H H OH H ⩽ ADP OH Glucose-6-phosphate The enzyme consists of two hinged lobes with the active site located between them (Fig. 6-14a). When glucose binds to hexokinase, the lobes swing together, engulfing the sugar (Fig. 6-14b). The result of hinge bending is that the substrate glucose is positioned near the substrate ATP such that a phosphoryl group can be easily transferred from the ATP to a hydroxyl group of the sugar. Not even a water molecule can enter the closed active site. This is beneficial, since water in the active site could lead to wasteful hydrolysis of ATP: ATP 1 H2O S ADP 1 Pi (b) Figure 6-14 Conformational changes in hexokinase. (a) The enzyme consists of two lobes connected by a hinge region. The active site is located in a cleft between the lobes. (b) When glucose (not shown) binds to the active site, the enzyme lobes swing together, enclosing glucose and preventing the entry of water. [Structure of “open” hexokinase (pdb 2YHX) determined by T. A. Steitz, C. M. Anderson, and R. E. Stenkamp; structure of “closed” hexokinase (pdb 1HKG) determined by W. S. Bennett Jr. and T. A. Steitz.] A glucose molecule in solution is surrounded by ordered water molecules in a hydration shell (see Section 2-1). These water molecules must be shed in order for glucose to fit into the active site of an enzyme, such as hexokinase. However, once the desolvated substrate is in the enzyme active site, the reaction can proceed quickly because there are no solvent molecules to interfere. In solution, rearranging the hydrogen bonds of surrounding water molecules as the reactants approach each other and pass through the transition state is energetically costly. By sequestering substrates in the active site, an enzyme can eliminate the energy barrier imposed by the ordered solvent molecules, thereby accelerating the reaction. This phenomenon is sometimes described as electrostatic catalysis since the nonaqueous active site allows more powerful electrostatic interactions between the enzyme and substrate than could occur in aqueous solution (for example, a low-barrier hydrogen bond can form in an active site but not in the presence of solvent molecules that would form ordinary hydrogen bonds). CONCEPT REVIEW • • • • What are the roles of proximity and orientation effects in enzyme catalysis? Why does the exclusion of water from an active site promote catalysis? Why is the lock-and-key model of enzyme action only partially valid? What is the role of the oxyanion hole in chymotrypsin? How does a low-barrier hydrogen bond promote catalysis? 6-4 Some Additional Features of Enzymes KEY CONCEPTS • Evolution has produced a number of serine proteases that differ in overall structure and substrate specificity. • Inactive zymogens are activated by proteolysis. • Active enzymes may be regulated by inhibitors. Chymotrypsin serves as a model for the structures and functions of a large family of serine proteases. And as with myoglobin and hemoglobin (Section 5-1), a close look at chymotrypsin reveals some general features of enzyme function, including evolution, substrate specificity, and inhibition. Not all serine proteases are related by evolution The first three proteases to be examined in detail were the digestive enzymes chymotrypsin, trypsin, and elastase, which have strikingly similar three-dimensional structures (Fig. 6-15). This was not expected on the basis of their limited sequence similarity (Table 6-3). However, careful examination revealed that most of the sequence variation is on the enzyme surface, and the positions of the catalytic residues in the three actives sites are virtually identical. It is believed that these proteins diverged from a common ancestor and have retained their overall structure and catalytic mechanism. Some bacterial proteases with a catalytically essential Ser are structurally related to the mammalian digestive serine proteases. However, the bacterial serine protease subtilisin (Fig. 6-16) shows no sequence similarity to chymotrypsin and no overall structural similarity, although it has the same Asp–His–Ser catalytic triad and an oxyanion hole in its active site. Subtilisin is an example of convergent evolution, a phenomenon whereby unrelated proteins evolve similar characteristics. As many as five groups of serine proteases, each with a different overall backbone conformation, have undergone convergent evolution to arrive at the same Asp, His, and Ser catalytic groups. In some other hydrolases, the substrate is attacked by a nucleophilic Ser or Thr residue that is located in a catalytic triad such as His–His–Ser or Asp–Lys–Thr. It would appear that natural selection favors this sort of arrangement of catalytic residues. Enzymes with similar mechanisms exhibit different substrate specificity Despite similarities in their catalytic mechanisms, chymotrypsin, trypsin, and elastase differ significantly from one another in their substrate specificity. Chymotrypsin preferentially cleaves peptide bonds following large hydrophobic residues. Trypsin prefers the basic residues Arg and Lys, and elastase cleaves the peptide bonds following small hydrophobic residues such as Ala, Gly, and Val (these residues predominate in elastin, an animal protein responsible for the elasticity of some tissues). The varying specificities of these enzymes are largely explained by the chemical character of the so-called specificity pocket, a cavity on the enzyme surface at the active site that accommodates the residue on the N-terminal side of the scissile peptide bond (Fig. 6-17). In chymotrypsin, the specificity pocket is about 10 Å deep and 5 Å wide, which offers a snug fit for an aromatic ring (whose dimensions are 6 Å 3 3.5 Å). The specificity pocket in trypsin is similarly sized but has an Asp residue rather than Ser at the bottom. Consequently, the trypsin specificity pocket readily binds the side chain of Arg or Lys, which has a diameter of about 4 Å and a cationic group at the Figure 6-15 Structures of chymotrypsin, trypsin, and elastase. The superimposed backbone traces of bovine chymotrypsin (blue), bovine trypsin (green), and porcine elastase (red) are shown along with the side chains of the active site Asp, His, and Ser residues. [Chymotrypsin structure (pdb 4CHA) determined by H. Tsukada and D. M. Blow; trypsin structure (pdb 3PTN) determined by J. Walker, W. Steigemann, T. P. Singh, H. Bartunik, W. Bode, and R. Huber; elastase structure (3EST) determined by E. F. Meyer, G. Cole, R. Radhakrishnan, and O. Epp.] Scissile bond Lys Ala Phe Gly 216 Gly 216 Gly 226 Gly 226 Ser 189 Chymotrypsin + − Thr 216 from Bacillus amyloliquefaciens. Val 226 The residues of the catalytic triad are highlighted in red. Compare the structure of this enzyme with those of the three serine proteases shown in Figure 6-15. Asp 189 Trypsin Figure 6-16 Structure of subtilisin Elastase [Structure (pdb 1CSE) determined by W. Bode.] Figure 6-17 Specificity pockets of three serine proteases. The side chains of key residues that determine the size and nature of the specificity pocket are shown along with a representative substrate for each enzyme. Chymotrypsin prefers large hydrophobic side chains; trypsin prefers Lys or Arg; and elastase prefers Ala, Gly, or Val. For convenience, the residues of all three enzymes are numbered to correspond to the sequence of residues in chymotrypsin. ? What would the specificity pocket look like in a protease that cleaved bonds following Asp or Glu residues? TABLE 6-3 Percent Sequence Identity among Three Serine Proteases Bovine trypsin Bovine chymotrypsinogen Porcine elastase 100% 53% 48% 176 Ch 6 How Enzymes Work end. In elastase, the specificity pocket is only a small depression due to the replacement of two Gly residues on the walls of the specificity pocket (residues 216 and 226 in chymotrypsin) with the bulkier Val and Thr. Elastase therefore preferentially binds small nonpolar side chains. Although these same side chains could easily enter the chymotrypsin and trypsin specificity pockets, they do not fit well enough to immobilize the substrate at the active site, as required for efficient catalysis. Other serine proteases, such as those that participate in blood coagulation, exhibit exquisite substrate specificity, in keeping with their highly specialized physiological functions. Chymotrypsin is activated by proteolysis Relatively nonspecific proteases could do considerable damage to the cells where they are synthesized unless their activity is carefully controlled. In many organisms, the activity of proteases is limited by the action of protease inhibitors (some of which are discussed further below) and by synthesizing the proteases as inactive precursors (called zymogens) that are later activated when and where they are needed. The inactive precursor of chymotrypsin is called chymotrypsinogen, and it is synthesized by the pancreas along with the zymogens of trypsin (trypsinogen), elastase (proelastase), and other hydrolytic enzymes. All these zymogens are activated by proteolysis after they are secreted into the small intestine. An intestinal protease called enteropeptidase activates trypsinogen by catalyzing the hydrolysis of its Lys 6OIle 7 bond. Enteropeptidase catalyzes a highly specific reaction; it appears to recognize a string of Asp residues near the N-terminus of its substrate: ⩽ H3N Val Asp Asp Asp H2O ⩽ H3N Val Asp Asp Asp Asp Lys Ile enteropeptidase Lys ⩽ COO⩾ ⩽ H3N Ile Trypsin, now itself active, cleaves the N-terminal peptide of the other pancreatic zymogens, including trypsinogen. The activation of trypsinogen by trypsin is an example of autoactivation. The Arg 15OIle 16 bond of chymotrypsinogen is susceptible to trypsin-catalyzed hydrolysis. Cleavage of this bond generates a species of active chymotrypsin (called p chymotrypsin), which then undergoes two autoactivation steps to generate fully active chymotrypsin (also called a chymotrypsin; Fig. 6-18). A similar process, in which zymogens are sequentially activated through proteolysis, occurs during blood coagulation (Box 6-B). Figure 6-18 Activation of chymotrypsinogen. Trypsin activates chymotrypsinogen by catalyzing hydrolysis of the Arg 15OIle 16 bond of the zymogen. The resulting active chymotrypsin then excises the Ser 14–Arg 15 dipeptide (by cleaving the Leu 13OSer 14 bond) and the Thr 147–Asn 148 dipeptide (by cleaving the Tyr 146OThr 147 and Asn 148OAla 149 bonds). All three species of chymotrypsin (p, d, and a) have proteolytic activity. 1 245 Chymotrypsinogen (inactive) 245 π Chymotrypsin (active) 245 δ Chymotrypsin (active) 245 α Chymotrypsin (active) trypsin 15 16 1 Arg Ile 14 15 chymotrypsin Ser Arg 1 13 16 Leu Ile 147 148 chymotrypsin Thr Asn 1 13 16 146 149 Leu Ile Tyr Ala BOX 6-B CLINICAL CONNECTION Blood Coagulation Requires a Cascade of Proteases When a blood vessel is injured by mechanical force, infection, or some other pathological process, red and white blood cells and the plasma (fluid) that surrounds them can leak out. Except in the most severe trauma, the loss of blood can be halted through formation of a clot at the site of injury. The clot consists of aggregated platelets (tiny enucleated cells that rapidly adhere to the damaged vessel wall and to each other) and a mesh of the protein fibrin, which reinforces the platelet plug and traps larger particles such as red blood cells. The conversion of fibrinogen to fibrin is the final step of coagulation, a series of proteolytic reactions involving a number of proteins and additional factors from platelets and damaged tissue. The enzyme responsible for cleaving fibrinogen to fibrin is known as thrombin. It is similar to trypsin in its sequence (38% of their residues are identical) and in its structure and catalytic mechanism. [Thrombin structure (pdb 1PPB) determined by W. Bode.] [P. Motta/Dept. of Anatomy, University La Sapienza, Rome/Science Photo Library/Photo Researchers.] Fibrin polymers can form rapidly because they are generated at the site of injury from the soluble protein fibrinogen, which circulates in the blood plasma. Fibrinogen is an elongated molecule with a molecular mass of 340,000 and consists of three pairs of polypeptide chains. The proteolytic removal of short (14- or 16-residue) peptides from the N-termini of four of the six chains causes the protein to polymerize in end-to-end and side-to-side fashion to produce a thick fiber. Fibrinogen thrombin Fibrin monomer Fibrin polymer Compare the structure of thrombin shown here with that of chymotrypsin (Fig. 6-1). Thrombin's catalytic residues (Asp, His, and Ser) are highlighted in red. Like trypsin, thrombin cleaves peptide bonds following Arg residues, but it is highly specific for the two cleavage sites in the fibrinogen sequence. Thrombin, like fibrinogen, circulates as an inactive precursor. Its zymogen, called prothrombin, contains a serine protease domain along with several other structural motifs. These elements interact with other coagulation factors to help ensure that thrombin—and therefore fibrin—is produced only when needed. A serine protease known as factor Xa catalyzes the specific hydrolysis of prothrombin to generate thrombin. Factor Xa (the a stands for active) is the protease form of the zymogen factor X. To initiate coagulation, factor X is activated by a protease known as factor VIIa, working in association with an accessory protein called tissue factor, which is exposed when a blood vessel is broken. During the later stages of coagulation, factor Xa is generated by the activity of factor IXa, which is generated from its zymogen by the activity of factor XIa or factor VIIa–tissue factor. Factor XIa is in turn generated by the proteolysis of its zymogen by trace amounts of thrombin produced earlier in the coagulation process. The cascade of activation reactions is depicted on the next page. Many of these enzymatic reactions require accessory factors that are not shown in this simplified diagram. 177 coagulation proteases. The protease recognizes the inhibitor as a substrate but is unable to complete the hydrolysis reaction. The protease and inhibitor form a stable acyl–enzyme intermediate that is removed from the circulation within a few minutes. Thrombin XI XIa or VIIa–tissue factor IX X IXa or VIIa–tissue factor * Xa Prothrombin Thrombin Fibrinogen Fibrin The coagulation cascade. The triggering step is marked with an asterisk. Note that the coagulation proteases are named according to their order of discovery, not their order of action (thrombin is also know as factor II). All the coagulation proteases appear to have evolved from a trypsin-like enzyme but have acquired narrow substrate specificity and a correspondingly narrow range of physiological activities. The coagulation reactions have an amplifying effect because each protease is a catalyst for the activation of another catalyst. Thus, a very small amount of factor IXa can activate a larger amount of factor Xa, which can then activate an even larger amount of thrombin. This amplification effect is reflected in the plasma concentrations of the coagulation factors. Plasma concentrations of some human coagulation factors Factor XI IX VII X Prothrombin Fibrinogen Concentration (mM)a 0.06 0.09 0.01 0.18 1.39 8.82 [Antithrombin structure (pdb 2ANT) determined by R. Skinner, J. P. Abrahams, J. C. Whisstock, A. M. Lesk, R. W. Carrell, and M. R. Wardell.] Heparin, a sulfated polysaccharide (Section 11-3) first purified from liver, enhances the activity of antithrombin by two mechanisms: A short segment (5 monosaccharide residues) acts as an allosteric activator of antithrombin, and a longer heparin polymer (containing at least 18 residues) can bind simultaneously to both antithrombin and its target protease such that their co-localization dramatically increases the rate of their reaction. Heparin or a synthetic version of it is used clinically as an anticoagulant following surgery. Defects in many of the proteins involved in blood coagulation or its regulation have been linked to bleeding (or clotting) disorders. For example, one form of hemophilia, a tendency to bleed following minor trauma, results from a genetic deficiency of factor IX. A deficiency of antithrombin results in an increased risk of clot formation in the veins. If dislodged, the clots may end up blocking an artery in the lungs or brain, with dire consequences. 1. a Concentrations calculated from data in High, K. A., and Roberts, H. R., eds., Molecular Basis of Thrombosis and Hemostasis, Marcel Dekker (1995). 2. A complex process such as coagulation is subject to regulation at a variety of points, including the activation and inhibition of the various proteases. Protease inhibitors account for about 10% of the protein that circulates in plasma. One inhibitor, known as antithrombin, blocks the proteolytic activities of factor IXa, factor Xa, and thrombin, thereby limiting the extent and duration of clot formation. Residues 377–400 of the 432-residue antithrombin form a loop (yellow in the figure above right) that extends away from the rest of the protein, and an arginine residue (red) at position 393 serves as a “bait” for the Arg-specific 3. 4. Questions: Hemophiliacs who are deficient in factor IX can be treated with injections of pure factor IX. Occasionally, a patient develops antibodies to this material, and it becomes ineffective. In such cases, the patient can be given factor VII instead. Explain why factor VII injections would be a useful treatment for a factor IX deficiency. A genetic defect in factor IX causes hemophilia, a serious disease. However, a genetic defect in factor XI may have no clinical symptoms. Explain this discrepancy in terms of the cascade mechanism for activation of coagulation proteases. Factor IX deficiency has been successfully treated by gene therapy, in which a normal copy of the gene is introduced into the body (see Section 3-4). Explain why restoration of factor IX levels to just a few percent of normal is enough to cure the abnormal bleeding. Of all the proteases included in the cascade diagram of coagulation, which one is the oldest in evolutionary terms? 5. Following severe trauma, surgery, or infection, a patient may develop disseminated intravascular coagulation (DIC), in which numerous small clots form throughout the circulatory system. Explain why patients with DIC subsequently exhibit excessive bleeding. 6. In one thrombin variant, a point mutation produces a protease with normal activity toward fibrinogen but decreased activity toward antithrombin. Would this genetic defect increase the risk of bleeding or of clotting? 7. Why can't heparin be administered orally? 8. Researchers have developed drugs based on hirudin, a thrombin inhibitor from the medicinal leech Hirudo medicinalis. Why would the leech be a good source for an anticoagulant? The two dipeptides that are excised during chymotrypsinogen activation are far removed from the active site (Fig. 6-19). How does their removal boost catalytic activity? A comparison of the X-ray structures of chymotrypsin and chymotrypsinogen reveals that the conformations of their active site Asp, His, and Ser residues are virtually identical (in fact, the zymogen can catalyze hydrolysis extremely slowly). However, the substrate specificity pocket and the oxyanion hole are incompletely formed in the zymogen. Proteolysis of the zymogen elicits small conformational changes that open up the substrate specificity pocket and oxyanion hole. Thus, the enzyme becomes maximally active only when it can efficiently bind its substrates and stabilize the transition state. Protease inhibitors limit protease activity The pancreas, in addition to synthesizing the zymogens of digestive proteases, synthesizes small proteins that act as protease inhibitors. The liver also produces a variety of protease inhibitor proteins that circulate in the bloodstream. If the pancreatic enzymes were prematurely activated or escaped from the pancreas through trauma, they would be rapidly inactivated by protease inhibitors. The inhibitors pose as protease substrates but are not completely hydrolyzed. For example, when trypsin attacks a Lys residue of bovine pancreatic trypsin inhibitor, the reaction halts during formation of the tetrahedral intermediate. The inhibitor remains in the active site, preventing any further catalytic activity (Fig. 6-20). The noncovalent complex between trypsin and bovine pancreatic trypsin inhibitor is one of the strongest protein– protein interactions known, with a dissociation constant of 10214 M. An imbalance between the activities of proteases and the activities of protease inhibitors may contribute to disease (see Box 6-B). Figure 6-19 Location of the dipeptides removed during the activation of chymotrypsinogen. The Ser 14–Arg 15 dipeptide (lower right, green) and the Thr 147–Asn 148 dipeptide (right, blue) are located at some distance from the active site residues (red) in chymotrypsinogen. [Chymotrypsinogen structure (pdb 2CGA) determined by D. Wang, W. Bode, and R. Huber.] Figure 6-20 The complex of trypsin with bovine pancreatic trypsin inhibitor. Ser 195 of trypsin (gold) attacks the peptide bond of Lys 15 (green) in the inhibitor, but the reaction is arrested on the way to the tetrahedral intermediate stage. [Structure (pdb 2PTC) determined by R. Huber and J. Deisenhofer.] CONCEPT REVIEW • • • • How are chymotrypsin, trypsin, and elastase similar? What is the chemical basis for their different substrate specificities? How is chymotrypsin activated and why is this step necessary? Why are protease inhibitors necessary? 179 180 SUMMARY 6-1 What Is an Enzyme? • Enzymes accelerate chemical reactions with high specificity under mild conditions. 6-2 The Chemistry of Catalysis • A reaction coordinate diagram illustrates the change in free energy between the reactants and products as well as the activation energy required to reach the transition state. The higher the activation energy, the fewer the reactant molecules that can reach the transition state and the slower the reaction. • An enzyme provides a route from reactants (substrates) to products, which has a lower activation energy than the uncatalyzed reaction. Enzymes, sometimes with the assistance of a cofactor, use chemical catalytic mechanisms such as acid–base catalysis, covalent catalysis, and metal ion catalysis. • In chymotrypsin, an Asp–His–Ser triad catalyzes peptide bond hydrolysis through acid–base and covalent catalysis and by stabilizing the transition state via the oxyanion hole and lowbarrier hydrogen bonds. 6-3 The Unique Properties of Enzyme Catalysts • In addition to transition state stabilization, enzymes use proximity and orientation effects, induced fit, and electrostatic catalysis to facilitate reactions. 6-4 Some Additional Features of Enzymes • Serine proteases that have evolved from a common ancestor share their overall structure and catalytic mechanism but differ in their substrate specificity. • The activities of some proteases are limited by their synthesis as zymogens that are later activated and by their interaction with protease inhibitors. GLOSSARY TERMS hydrolysis reactant catalyst enzyme ribozyme active site reaction specificity substrate isozymes reaction coordinate DG ‡ transition state DGreaction cofactor coenzyme acid–base catalysis acid catalysis base catalysis tautomer carbanion covalent catalysis imine Schiff base nucleophile electrophile metal ion catalysis chemical labeling serine protease catalytic triad scissile bond lock-and-key model oxyanion hole low-barrier hydrogen bond proximity and orientation effects induced fit electrostatic catalysis convergent evolution specificity pocket zymogen autoactivation coagulation protease inhibitor PROBLEMS 6-1 What Is an Enzyme? 1. Why are most enzymes globular rather than fibrous proteins? 2. Explain why the motor proteins myosin and kinesin (described in Section 5-3) are enzymes and write the reaction that each catalyzes. 3. The uncatalyzed rate of amide bond hydrolysis in hippurylphenylalanine is 1.3 3 10210 s21, whereas the rate of hydrolysis of the amide bond in the same substrate catalyzed by carboxypeptidase is 61 s21. What is the rate enhancement for amide bond hydrolysis by this enzyme for this particular substrate? 4. The half-life for the hydrolysis of the glycosidic bond in the sugar trehalose is 6.6 3 106 years. (a) What is the rate constant for the uncatalyzed hydrolysis of this bond? (Hint: For a first-order reaction, the rate constant, k, is equal to 0.693/t1/2.) (b) What is the rate enhancement for glycosidic bond hydrolysis catalyzed by trehalase if the rate constant for the catalyzed reaction is 2.6 3 103 s21? 5. Compare the rate enhancements of adenosine deaminase and triose phosphate isomerase (see Table 6-1). 6. What is the relationship between the rate of an enzymecatalyzed reaction and the rate of the corresponding uncatalyzed reaction? Do enzymes enhance the rates of slow uncatalyzed reactions as much as they enhance the rates of fast uncatalyzed reactions? 7. Draw an arrow to identify the bonds in the peptide below that would be hydrolyzed in the presence of the appropriate peptidase. ⩽ H N N H3N O O O N H H N O O O⩾ 181 8. (a) The molecule shown here is a substrate for trypsin. Draw the reaction products. 11. Draw the oxidized product of the reaction catalyzed by succinate dehydrogenase. To which class of enzymes does succinate dehydrogenase belong? NO2 COO⩾ H C H O H C H NH HN COO⩾ Succinate O 12. Malate dehydrogenase catalyzes a reaction in which C2 of malate is oxidized. Draw the structure of the product. To which class of enzymes does malate dehydrogenase belong? HN NH2 ⩽ succinate dehydrogenase H2N 9. The reactions catalyzed by the enzymes listed below are presented in this chapter. To which class does each enzyme belong? Explain your answers. (a) pyruvate decarboxylase (b) alanine aminotransferase (c) alcohol dehydrogenase (d) hexokinase (e) chymotrypsin COO⩾ CH2 (b) Design an assay that would allow you to measure the rate of trypsin-catalyzed hydrolysis of the molecule shown in part (a). OH CH COO⩾ Malate 13. Examine the reaction catalyzed by hexokinase (Section 6-3). Draw the product of the reaction catalyzed by creatine kinase, which acts on creatine in a similar manner. What would you predict to be the usual function of a kinase? NH2 10. To which class do the enzymes that catalyze the following reactions belong? (a) O C H CH OH CH2 O (b) COO⩾ CH O PO32⩾ HO N CH3 C CH2COO⩾ O O Creatine PO2⩾ 3 14. Propose a name for the enzymes that catalyze the following reactions (reactions may not be balanced): C O PO32⩾ O HO C H HO H H CH3 CH2OH O OH H H HPO42⩾ H HO PO32⩾ HO H OH COO⩾ CH2 CH COO⩾ CH COO⩾ Isocitrate OH H HO CH2OH O OH H H OH (c) H OH HC OH CH2 COO⩾ CH2 ⩾ COO ⩽ O C H Succinate OH CH2OPO32⩾ 1,3-Bisphosphoglycerate ADP COO⩾ Glyoxylate O ATP OPO32⩾ O 6-Phosphoglucono--lactone O C CH2OPO32⩾ O H OH H H Glucose-6-phosphate (b) O CH2OPO32⩾ O OH H OH H H (a) COO⩾ CH3 H CH2OH CH2 (c) COO⩾ (d) NH⩽ 2 COO⩾ CH2OH C C CH2 PO2⩾ 3 malate dehydrogenase C O⩾ HC OH CH2OPO32⩾ 3-Phosphoglycerate 182 (d) CH3 C CH2 C O COO⩾ CO2 COO⩾ H HO O CH2 1 Oxaloacetate HO 15. The amino acid tryptophan is converted to the hormone melatonin as shown here. Indicate which reaction is catalyzed by each of the following types of enzymes. (a) methyltransferase (b) hydroxylase (c) acetyltransferase (d) decarboxylase H HO CH2 2 HO CH2 HO HO NH⩽ 3 C HO HO NH⩽ 3 O CH3 6-2 The Chemistry of Catalysis N 19. Approximately how much does staphylococcal nuclease (Table 6-1) decrease the activation energy (DG ‡) of its reaction (the hydrolysis of a phosphodiester bond)? H O N H N H 18. What is the common name of the enzyme whose EC number is 4.3.2.1? NH⩽ 3 4 C 17. Use the enzyme nomenclature database at . expasy.org to determine what reaction is catalyzed by the enzyme called catalase. H HO CH2 OH H N 3 NH⩽ 3 H O⩾ HO CH2 4 O HO 2 NH⩽ 3 CH2 OH HO H N COO⩾ NH⩽ 3 O⩾ 1 C 3 H N COO⩾ NH⩽ 3 COO⩾ Pyruvate C CH3 H 20. The rate of an enzyme-catalyzed reaction is measured at several temperatures, generating the curve shown below. Explain why the enzyme activity increases with temperature and then drops off sharply. N O H3CO N CH3 H Enzyme activity 16. The amino acid tyrosine is converted to the hormone epinephrine as shown above right. Indicate which reaction is catalyzed by each of the following types of enzymes. (a) decarboxylase (b) methyltransferase (c) hydroxylase 20 40 60 Temperature (°C) 80 183 21. Draw pairs of free energy diagrams that show the differences between the following reactions: (a) a fast reaction vs. a slow reaction (b) a one-step reaction vs. a two-step reaction (c) an endergonic reaction vs. an exergonic reaction (d) a two-step reaction with an initial slow step vs. a two-step reaction with an initial fast step 22. Consult Figure 6-10 and draw a reaction coordinate diagram for the chymotrypsin-catalyzed hydrolysis of a peptide bond. 23. Under certain conditions, peptide bond formation is more thermodynamically favorable than peptide bond hydrolysis. Would you expect chymotrypsin to catalyze peptide bond formation? transmission. Parathion doesn't react with the enzyme directly; it is first converted to paraoxon, which then reacts with acetylcholinesterase in a manner similar to the reaction of DIPF with chymotrypsin. Draw the structure of the enzyme's catalytic serine reside modified by paraoxon. OCH2CH3 O 2N O P S OCH2CH3 Parathion OCH2CH3 24. What is the relationship between the nucleophilicity and the acidity of an amino acid side chain? 25. Amino acids such as Gly, Ala, and Val are not known to participate directly in acid–base or covalent catalysis. (a) Explain why this is the case. (b) Mutating a Gly, Ala, or Val residue in an enzyme's active site can still have dramatic effects on catalysis. Why? 26. Using what you know about the structures of the amino acid side chains and the mechanisms presented in this chapter, choose an amino acid side chain to play the following roles in an enzymatic mechanism: (a) participate in proton transfer, (b) act as a nucleophile. 27. Ribozymes are RNA molecules that catalyze chemical reactions. (a) What features of a nucleic acid would be important for it to act as an enzyme? (b) What part of the RNA structure might fulfill the roles listed in Problem 26? (c) Why can RNA but not DNA act as a catalyst? 28. RNA is susceptible to base catalysis in which the hydroxide ion abstracts a proton from the 29 OH group and then the resulting 29 O2 nucleophilically attacks the 59 phosphate. (a) Is DNA susceptible to base hydrolysis? (b) Does this observation help you explain why DNA rather than RNA evolved as the genetic material? 29. Use what you know about the mechanism of chymotrypsin to explain why DIPF inactivates the enzyme. 30. Solutions of chymotrypsin prepared for experimental purposes are stored in a dilute solution of hydrochloric acid as a “preservative.” Explain. 31. Sarin is an organophosphorous compound similar to DIPF. In 1995, terrorists released it on a Japanese subway. Injuries resulted from the reaction of sarin with a serine esterase involved in nerve transmission called acetylcholinesterase. Draw the structure of the enzyme's catalytic residue modified by sarin. H3C H3C O P O P O OCH2CH3 Paraoxon 33. Angiogenin, an enzyme involved in the formation of blood vessels, was treated with bromoacetate (BrCH2COO2). An essential histidine was modified, and the activity of the enzyme was reduced by 95%. Show the chemical reaction for the modification of the histidine side chain with bromoacetate. Why would such a modification inactivate an enzyme that required a His side chain for activity? 34. (a) Diagram the hydrogen-bonding interactions of the catalytic triad His–Lys–Ser during catalysis in a hypothetical hydrolytic enzyme. (b) In a chemical labeling study of the enzyme, a reagent that covalently modifies Lys residues also abolishes enzyme activity. Why doesn't this observation prove that the active site includes a Lys residue? 35. Treatment of the enzyme D-amino acid oxidase with 1-fluoro2,4-dinitrophenol (FDNP, below) inactivates the enzyme. Analysis of the derivatized enzyme revealed that one of the tyrosines in the enzyme was unusually reactive. In a second experiment, when benzoate (a compound that structurally resembles the enzyme's substrate) was added prior to the addition of the FDNP, the tyrosine did not react. (a) Show the chemical reaction between FDNP and the tyrosine residue. (b) Why does benzoate prevent the reaction? What does this tell you about the role of the tyrosine in this enzyme? F NO2 NO2 36. Propose two variants of a protease catalytic triad that take the form X–His–Y. O CH O 2N F CH3 Sarin 32. Parathion is an insecticide that kills pests by reacting irreversibly with acetylcholinesterase, a serine esterase involved in nerve 37. How would chymotrypsin's catalytic triad be affected by extremely low and extremely high pH values (assuming that the rest of the protein structure remained intact)? 38. E. coli ribonuclease HI is an enzyme that catalyzes the hydrolysis of phosphodiester bonds in RNA. Its proposed mechanism involves a “carboxylate relay.” 184 (a) Using the structures below as a guide, draw arrows that indicate how the hydrolysis reaction might be initiated. His 124 Asp 70 RNA Substrate O HN O CH C HN CH C HN N O C O O H H O O ⩾ O P O H O H Asp 52 ⩾OOC COOH O Asp 52 ⩾OOC Glu 35 ⩽ (a) Assign the given pK values to Glu 35 and Asp 52. (b) Lysozyme is inactive at pH 2.0 and at pH 8.0. Explain. (c) The lysozyme mechanism proceeds via a covalent intermediate. Use this information to complete the drawing of the lysozyme mechanism. 40. RNase A is a digestive enzyme secreted by the pancreas into the small intestine, where it hydrolyzes RNA into its component nucleotides. The optimum pH for RNase A is about 6, and the pK values of the two histidines that serve as catalytic residues are 5.4 and 6.4. The first step of the mechanism is shown. (a) Does ribonuclease proceed via acid–base catalysis, covalent catalysis, metal ion catalysis, or some combination of these strategies? Explain. (b) Assign the appropriate pK values to His 12 and His 119. (c) Explain why the pH optimum of ribonuclease is pH 6. ⩾ O CH2 H NH H P N His 12 2⬘ H O O CH2 O HN N CH2 Base H ⩽ O H H H OH O H ⩾ O Base P O O 41. Chymotrypsin hydrolyzes the artificial substrate p-nitrophenylacetate using a similar mechanism to that of peptide bond hydrolysis. The product p-nitrophenolate is bright yellow and absorbs light at 410 nm; thus the reaction can be monitored spectrophotometrically. When p-nitrophenylacetate is mixed with chymotrypsin, p-nitrophenolate initially forms extremely rapidly. This is followed by a steady-state phase in which the product forms at a uniform rate. Absorbance at 410 nm 39. Lysozyme catalyzes the hydrolysis of a polysaccharide component of bacterial cell walls. The damaged bacteria subsequently lyse (rupture). Part of lysozyme's mechanism is shown below. The enzyme catalyzes cleavage of a bond between two sugar residues (represented by hexagons). Catalysis involves the side chains of Glu 35 and Asp 52. One of the residues has a pK of 4.5; the other has a pK of 5.9. HO H O (b) The pK values for all of the histidines in ribonuclease HI were determined and include values of 7.1, 5.5, and ,5.0. Which value is most likely to correspond to His 124? Explain. (c) Substituting an alanine residue for His 124 resulted in a dramatic decrease in enzyme activity. Explain. COO⩾ O 3⬘ H His 119 Glu 35 CH2 O O O H P O⩾ CH2 CH2 (d) Ribonuclease catalyzes the hydrolysis of RNA but not DNA. Explain why. Time (a) Using what you know about the mechanism of chymotrypsin, explain why these results were observed. (b) Draw a reaction coordinate diagram for the reaction. (c) Would p-nitrophenylacetate be useful for monitoring the activity of trypsin in vitro? 42. Renin is similar to chymotrypsin except that it belongs to a class of enzymes called aspartyl proteases in which aspartate replaces serine in the enzyme active site. Renin inhibitors are potentially valuable in the treatment of high blood pressure. Renin catalyzes the hydrolysis of a peptide bond in the protein angiotensinogen, converting it to angiotensin I, a precursor of a hormone involved in regulating blood pressure. Two aspartate residues in the renin active 185 site, Asp 32 and Asp 215, constitute a catalytic dyad, rather than the catalytic triad found in chymotrypsin. (a) Begin with the figure provided and draw the reaction mechanism for peptide hydrolysis catalyzed by renin. (b) What type of catalytic strategy is utilized by renin? (c) Compare the pK values of the Asp 32 and Asp 215 residues. enzyme belongs to a family of enzymes called metalloproteases. The active site includes a Zn21 ion coordinated by two His residues and a Glu residue. The active site contains a second Glu residue (Glu 224). Use the figure provided as a starting point to draw the mechanism of peptide bond hydrolysis catalyzed by this enzyme. H O Asp 32 C O O⩾ H H C O⩾ Glu 224 H H O O N C O O Zn2⩽ H N C OH C (CH2)2 His His Glu O Asp 215 43. The protease from the human immunodeficiency virus (HIV) is a target of the “cocktail” of drugs now on the market to treat HIV/AIDS. The protease has a mechanism similar to that of renin, described in Problem 42, except that renin consists of a single protein with two catalytic Asp residues, whereas the HIV protease consists of two identical subunits, each of which contributes an Asp residue. Is this an example of convergent or divergent evolution? Explain. 44. The enzyme bromelain (found in pineapple) is also a protease but is a member of a family of enzymes, referred to as cysteine proteases, in which a cysteine residue plays a catalytic role similar to serine in serine proteases. Bromelain has been found to relieve inflammation, and some studies have shown that it has antitumor activity. Like the serine proteases, bromelain contains a histidine, but unlike serine proteases, bromelain lacks an aspartate in its active site. (a) Using what you know about chymotrypsin's mechanism, draw a reaction mechanism for the hydrolysis of a peptide bond by bromelain. (b) Does the mechanism you have drawn employ acid–base catalysis, covalent catalysis, or both? (c) Draw a reaction coordinate diagram that is consistent with the mechanism of bromelain. (d) Why did natives of tropical countries use bromelain as a meat tenderizer? (e) The pK values for the active site residues in bromelain are ,3 and ,8. The pH optimum for the reaction is 6.0. Assign pK values to the appropriate amino acid side chains and explain your reasoning. 45. Sulfhydryl groups can react with the alkylating reagent N-ethylmaleimide (NEM). When NEM is added to a solution of creatine kinase, Cys 278 is alkylated, but no other Cys residues in the protein are modified. What can you infer about the role of the Cys 278 residue based on this information? 46. Cysteine proteases (see Problem 44) can often be inactivated by iodoacetate (ICH2COO2). Show the chemical reaction for the modification of the Cys side chain by

iodoacetate. Why would such a modification inactivate an enzyme that required a Cys side chain for activity? 47. The botulinum neurotoxin secreted by Clostridium botulinum consists of three polypeptide chains, one of which is a protease. The 48. The enzyme carbonic anhydrase is one of the fastest enzymes known and, like the enzyme in Problem 47, is a metalloenzyme. It catalyzes the hydration of carbon dioxide to form bicarbonate and a hydrogen ion: CO2 1 H2O Δ H 1 1 HCO32 The enzyme's active site contains a zinc ion that is coordinated to the imidazole rings of three histidine residues. (A fourth histidine residue is located nearby and participates in catalysis). A fourth coordination position is occupied by a water molecule. Draw the reaction mechanism (the first step is shown) of the hydration of carbon dioxide and show the regeneration of the enzyme. O HN CH C His H CH2 O HN H N Zn2⩽ His His 49. Asn and Gln residues in proteins are sometimes nonenzymatically hydrolytically deamidated to Asp and Glu residues, respectively. It has been suggested that deamidation might function as a molecular timer for protein turnover, since deamidation of some proteins increases their susceptibility to degradation by cellular proteases. An exhaustive study of proteins known to undergo deamidation has revealed that deamidated Asn residues are most likely to be preceded by Ser, Thr, or Lys and to be followed by Gly, Ser, or Thr. (a) Write the balanced chemical equation for the hydrolytic deamidation of Asn to Asp. (b) The mechanism of deamidation of the amide side chain involves an acid catalyst (HA). The first step of the reaction is shown below. Draw the rest of the reaction mechanism. CH2 CH2 C O NH2 H A 186 (c) Since the deamidation of Asn residues is known to be nonenzymatic, the HA acid catalyst cannot be provided by an enzyme. Instead, the catalytic groups are believed to be provided by neighboring amino acids in the protein undergoing deamidation. Refer to your answer in part (b) and examine the mechanism you have written. Describe how amino acid side chains that either precede or follow the labile Asn could serve as catalytic groups in the deamidation process. (d) Amino terminal Gln residues undergo deamidation much more rapidly than internal Gln residues. Write a mechanism for this deamidation process, which includes the formation of a fivemembered pyrrolidone ring. Amino terminal Asn residues do not undergo deamidation. Explain why. (e) It has been observed that Asn and Gln residues in the interior of a protein are deamidated at a much slower rate than Asn and Gln residues on the surface of the protein. Explain why. NH2 N 1 O 6 N N N adenosine deaminase N HN N Ribose H2O N NH3 Ribose Adenosine Inosine H H N HN N N Ribose 1,6-Dihydropurine nucleoside 50. Enzymes in the amidase family catalyze the hydrolysis of 58. Imidazole reacts with p-nitrophenylacetate to produce the amide bonds using a mechanism similar to that of the serine proteN-acetylimidazolium and p-nitrophenolate ions. ases. But the catalytic triad contains the residues Ser–Ser–Lys instead of the Asp–His–Ser triad found in the serine proteases. The Ser– O Ser–Lys triad is shown below. Propose a mechanism for the amidase enzyme. NO2 H3C C O CH2 O⩾ R2 C CH2 (CH2)4 HO H N NH⩽ 3 N p -Nitrophenylacetate O N H H3C Imidazole R1 O 6-3 The Unique Properties of Enzyme Catalysts 51. When substrates bind to an enzyme, their free energy may be lowered. Why doesn't this binding defeat catalysis? 52. Substrates and reactive groups in an enzyme's active site must be precisely aligned in order for a productive reaction to occur. Why, then, is some conformational flexibility also a requirement for catalysis? 53. Daniel Koshland has noted that hexokinase undergoes a large conformational change on substrate binding, which apparently prevents water from entering the active site and participating in hydrolysis. However, the serine proteases do not undergo large conformational changes on substrate binding. Explain. 54. When ATP and the sugar xylose (which resembles glucose but has only five carbons) are added to hexokinase, the enzyme produces xylose-5-phosphate along with some free phosphate. Does this observation support Koshland's induced fit hypothesis? 55. Refer to Problem 47. What is the role of the zinc ion in catalysis? In transition state stabilization? 56. Refer to Problem 50. How might the amidase stabilize the transition state? 57. The enzyme adenosine deaminase catalyzes the conversion of adenosine to inosine (above, right). The compound 1,6-dihydropurine ribonucleoside binds to the enzyme with a much greater affinity than the adenosine substrate. What does this tell us about the mechanism of adenosine deaminase? C N⩽ ⩽ ⩾O NO2 N-Acetylimidazolium N H p -Nitrophenolate (a) The compound shown below reacts to form p-nitrophenolate. Draw the reaction mechanism for this reaction. O C O NO2 N NH (b) The compound described in part (a) reacts 24 times faster than imidazole. Explain why. (c) What can the results of this experiment tell us about the way enzymes speed up reaction rates? 6-4 Some Additional Features of Enzymes 59. Many genetic mutations prevent the synthesis of a protein or give rise to an enzyme with diminished catalytic activity. Is it possible for a mutation to increase the catalytic activity of an enzyme? 60. If the active site of chymotrypsin excludes water when its substrate binds, how is it possible for a water molecule to participate in the second stage of catalysis (as shown in Fig. 6-10)? 61. The amino acid Asp 189 lies at the base of the substrate specificity pocket in the enzyme trypsin. 187 (a) How is this related to trypsin's substrate specificity? What kinds of interactions take place between the Asp 189 and the side chain of the residue preceding the scissile bond? (b) In site-directed mutagenesis studies, Asp 189 was replaced with lysine. How do you think this would affect substrate specificity? (c) The investigators who carried out the experiment described in part (b) analyzed the three-dimensional structure of the mutant enzyme and found that Lys 189 is actually not located in the substrate specificity pocket. Instead, the Lys side chain reaches out of the base of the pocket, rendering the specificity pocket nonpolar. With this additional information, determine how the substrate specificity would differ in the Lys 189 mutant enzyme. 62. An enzyme involved in the regulation of blood pressure was recently purified and characterized. It is an aspartyl aminopeptidase, which hydrolyzes peptide bonds on the carboxyl side of aspartate residues. The investigators who purified the enzyme were interested in learning about the substrate binding site, so they synthesized a series of artificial peptides and tested the ability of the enzyme to hydrolyze them. The peptides and their reaction rates are listed in the table above right. The residues are labeled P1–P19–P29–P39, and the hydrolyzed bond is between P1 and P19. (a) It is likely that both the P19 residue and the P29 residue fit in adjacent "pockets" on the aspartyl aminopeptidase enzyme. Describe the characteristics of these pockets on the enzyme, using the kinetic data in the table. (b) The investigators have proposed that the "natural" or endogenous substrate for aspartyl aminopeptidase is angiotensin II, which is involved in the regulation of blood pressure. However, the authors also mention that a potential exogenous substrate might be the artificial sweetener aspartame (Nutrasweet®), which is the methylated dipeptide aspartylphenylalanine (Asp–Phe– OCH3). Based on the results presented here, do you think that aspartame would be a good substrate for the enzyme? (c) Draw the structure of aspartame and the products of its hydrolysis by aspartyl aminopeptidase. (d) Under nondenaturing conditions, the enzyme appears to be the same size as the protein ferritin, which has a molecular mass of 440 kD. Under denaturing conditions, the purified enzyme behaves like a single polypeptide of 55 kD. What can you determine about the enzyme's structure from these data? Peptide (P1–P19–P29–P39) Catalyzed rate (s–1) Asp–Ala–Ala–Leu Asp–Phe–Ala–Leu Asp–Lys–Ala–Leu Asp–Asp–Ala–Leu Asp–Ala–Phe–Leu Asp–Ala–Lys–Leu Asp–Ala–Asp–Leu 5.3 9.9 2.8 9.8 17.2 5.0 2.3 63. Chymotrypsin is usually described as an enzyme that catalyzes hydrolysis of peptide bonds following Phe, Trp, or Tyr residues. Is this information consistent with the description of the bonds cleaved during chymotrypsin activation (Figure 6-18)? What does this tell you about chymotrypsin's substrate specificity? 64. Describe how the activation of the digestive enzyme chymotrypsin follows a cascade mechanism. Draw a diagram of your cascade. 65. Would chymotrypsin catalyze hydrolysis of the substrate shown in Problem 8? Explain why or why not. 66. Does DIPF inactivate trypsin or elastase (see Problem 29)? Explain. 67. Hereditary pancreatitis is caused by a mutation in the autocatalytic domain of trypsinogen that results in persistent activity of the enzyme. (a) What are the physiological consequences of this disease? (b) Describe a strategy to treat this disease. 68. Some plants contain compounds that inhibit serine proteases. It has been hypothesized that these compounds protect the plant from proteolytic enzymes of insects and microorganisms that would damage the plant. Tofu, or bean curd, possesses these compounds. Manufacturers of tofu treat it to eliminate serine protease inhibitors. Why is this treatment necessary? 69. Not all proteases are synthesized as zymogens or have inhibitors that block their activity. What limits the potentially destructive power of these proteases? 70. A chymotrypsin inhibitor isolated from snake venom resembles the bovine pancreatic trypsin inhibitor (Fig. 6-20) but has an Asn residue in place of Lys 15. Why is this finding unexpected? SELECTED READINGS Di Cera, E., Serine proteases, IUBMB Life 61, 510–515 (2009). [Summarizes some of the biological roles and mechanistic features of proteases.] Fersht, A., Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding, W. H. Freeman (1999). [Includes detailed reaction mechanisms for chymotrypsin and other enzymes.] Gutteridge, A., and Thornton, J. M., Understanding nature's catalytic toolkit, Trends Biochem. Sci. 30, 622–629 (2005). [Describes how certain sets of amino acid side chains form catalytic units that appear in many different enzymes.] Radisky, E. S., Lee, J. M., Lu, C.-J. K., and Koshland, D. E., Jr., Insights into the serine protease mechanism from atomic resolution structures of trypsin reaction intermediates, Proc. Nat. Acad. Sci. 103, 6835–6840 (2006). Ringe, D., and Petsko, G. A., How enzymes work, Science 320, 1428–1429 (2008). [Briefly summarizes some general features of enzyme function.] chapter 7 ENZYME KINETICS AND INHIBITION HOW do we measure how well a drug works? One of the best-selling pharmaceutical compounds is atorvastatin (Lipitor, shown here in space-filling form). Like many other drugs, atorvastatin inhibits the catalytic activity of an enzyme, in this case HMG-CoA reductase, which catalyzes a critical step in the biosynthesis of cholesterol. Consequently, atorvastatin can lower the level of circulating cholesterol, thereby reducing the risk of cardiovascular disease. But long before the first patient with high cholesterol got a prescription for Lipitor, researchers had to characterize the drug's effects on the enzyme by measuring catalytic activity in the presence and absence of the drug, using the universal language of enzyme kinetics. THIS CHAPTER IN CONTEXT Do You Remember? Part 1 Foundations Living organisms obey the laws of thermodynamics (Section 1-3). Noncovalent forces, including hydrogen bonds, ionic interactions, and van der Waals forces, act on biological molecules (Section 2-1). Part 2 Molecular Structure and Function 7 Enzyme Kinetics and Inhibition Part 3 Metabolism Part 4 Genetic Information 188 An acid's pK value describes its tendency to ionize (Section 2-3). The 20 amino acids differ in the chemical characteristics of their R groups (Section 4-1). Some proteins can adopt more than one stable conformation (Section 4-3). In the preceding chapter, we examined the basic features of enzyme catalytic activity, primarily by exploring the various catalytic mechanisms used by chymotrypsin. This chapter extends the discussion of enzymes by introducing enzyme kinetics, the mathematical analysis of enzyme activity. Here we describe how an enzyme's reaction speed and specificity can be quantified and how this information can be used to evaluate the enzyme's physiological function. We also look at the regulation of enzyme activity by inhibitors that bind to the enzyme and alter its activity. The discussion also focuses on allosteric regulation, a mechanism for inhibiting as well as activating enzymes. 189 Introduction to Enzyme Kinetics 7-1 Introduction to Enzyme Kinetics The structure and chemical mechanism of an enzyme (for example, chymotrypsin, as discussed in Chapter 6) often reveal a great deal about how that enzyme functions in vivo. However, structural information alone does not provide a full accounting of an enzyme's physiological role. For example, one might need to know exactly how fast the enzyme catalyzes a reaction or how well it recognizes different substrates or how its activity is affected by other substances. These questions are not trivial— consider that a single cell contains thousands of different enzymes, all operating simultaneously and in the presence of one anothers' substrates and products. To fully describe enzyme activity, enzymologists apply mathematical tools to quantify an enzyme's catalytic power and its substrate affinity as well as its response to inhibitors. This analysis is part of the area of study known as enzyme kinetics (from the Greek kinetos, which means "moving"). Some of the earliest biochemical studies examined the reactions of crude preparations of yeast cells and other organisms. Even without isolating any enzymes, researchers could mathematically analyze their activity by measuring the concentrations of substrates and reaction products and observing how these quantities changed over time. For example, consider the simple reaction catalyzed by triose phosphate isomerase, which interconverts two three-carbon sugars (trioses): H H O H C C H C OH KEY CONCEPT • An enzyme's activity, measured as the rate of product formation, varies with the substrate concentration. triose phosphate isomerase C OH O CH2OPO32⩾ CH2OPO32⩾ Glyceraldehyde-3-phosphate Dihydroxyacetone phosphate Over the course of the reaction, the concentration of the substrate glyceraldehyde-3 phosphate falls as the concentration of the product dihydroxyacetone phosphate rises (Fig. 7-1). The progress of this or any reaction can be expressed as a velocity Figure 7-1 Progress of the triose phosphate isomerase reaction. Over Glyceraldehyde-3-phosphate (substrate) time, the concentration of the substrate glyceraldehyde-3-phosphate decreases and the concentration of the product dihydroxyacetone phosphate increases. Concentration ? Dihydroxyacetone phosphate (product) Time Extend the lines in the graph to show that the reaction eventually reaches equilibrium, when the ratio of product concentration to reactant concentration is about 22. 190 Ch 7 Enzyme Kinetics and Inhibition (v), either the rate of disappearance of the substrate (S) or the rate of appearance of the product (P): v52 Reaction velocity, v Enzyme concentration, [E] [7-1] where [S] and [P] represent the concentrations of the substrate and product, respectively. Not surprisingly, the more catalyst (enzyme) present, the faster the reaction (Fig. 7-2). When the enzyme concentration is held constant, the reaction velocity varies with the substrate concentration, but in a nonlinear fashion (Fig. 7-3). The shape of this velocity versus substrate curve is an important key to understanding how enzymes interact with their substrates. The hyperbolic, rather than linear, shape of the curve suggests that an enzyme physically combines with its substrate to form an enzyme–substrate (ES) complex. Therefore, the enzyme-catalyzed conversion of S to P Figure 7-2 Progress of the triose E S ¡ P phosphate isomerase reaction. The more enzyme present, the faster the reaction. d [S] d [P] 5 dt dt can be more accurately written as E 1 S S ES S E 1 P As small amounts of substrate are added to the enzyme preparation, enzyme activity (measured as the reaction velocity) appears to increase almost linearly. However, the enzyme's activity increases less dramatically as more substrate is added. At very high substrate concentrations, enzyme activity appears to level off as it approaches a maximum value. This behavior shows that at low substrate concentrations, the enzyme quickly converts all the substrate to product, but as more substrate is added, the enzyme becomes saturated with substrate—that is, there are many more substrate molecules than enzyme molecules, so not all the substrate can be converted to product in a given time. These so-called saturation kinetics are a feature of many binding phenomena, including the binding of O2 to myoglobin (see Section 5-1). The curve shown in Figure 7-3 reveals considerable information about a given enzyme and substrate under a chosen set of reaction conditions. All simple enzymecatalyzed reactions yield a hyperbolic velocity versus substrate curve, but the exact shape of the curve depends on the enzyme, its concentration, the concentrations of Reaction velocity, v Substrate concentration, [S] Figure 7-3 A plot of reaction velocity versus substrate concentration. Varying amounts of substrate are added to a fixed amount of enzyme. The reaction velocity is measured for each substrate concentration and plotted. The resulting curve takes the form of a hyperbola, a mathematical function in which the values first increase steeply but eventually approach a maximum. ? Compare this diagram to Figure 5-3, which shows oxygen binding to myoglobin. enzyme inhibitors, the pH, the temperature, and so on. By analyzing such curves, it is possible to address some basic questions, for example, 191 Derivation and Meaning of the Michaelis–Menten Equation • How fast does the enzyme operate? • How efficiently does the enzyme convert different substrates to products? • How susceptible is the enzyme to various inhibitors and how do these inhibitors affect enzyme activity? The answers to these questions, in turn, may reveal • Whether an enzyme is likely to catalyze a particular reaction in vivo • What substances are likely to serve as physiological regulators of the enzyme's activity • Which enzyme inhibitors might be effective drugs CONCEPT REVIEW • Describe two ways to express the velocity of an enzymatic reaction. • Why does a plot of velocity versus substrate concentration yield a curve? 7-2 Derivation and Meaning of the Michaelis–Menten Equation The mathematical analysis of enzyme behavior centers on the equation that describes the hyperbolic shape of the velocity versus substrate plot (see Fig. 7-3). We can analyze an enzyme-catalyzed reaction, such as the one catalyzed by triose phosphate isomerase, by conceptually breaking it down into smaller steps and using the terms that apply to simple chemical processes. Rate equations describe chemical processes Consider a unimolecular reaction (one that involves a single reactant) such as the conversion of compound A to compound B: ASB The progress of this reaction can be mathematically described by a rate equation in which the reaction rate (the velocity) is expressed in terms of a constant (the rate constant) and the reactant concentration [A]: v52 d [A] 5 k [A] dt [7-2] Here, k is the rate constant and has units of reciprocal seconds (s21). This equation shows that the reaction velocity is directly proportional to the concentration of reactant A. Such a reaction is said to be first-order because its rate depends on the concentration of one substance. A bimolecular or second-order reaction, which involves two reactants, can be written A1BSC Its rate equation is v52 d [A] d [B] 52 5 k[A][B] dt dt [7-3] Here, k is a second-order rate constant and has units of M21 ? s21. The velocity of a second-order reaction is therefore proportional to the product of the two reactant concentrations (see Sample Calculation 7-1). KEY CONCEPTS • Simple chemical reactions are described in terms of rate constants. • The Michaelis–Menten equation describes enzyme-catalyzed reactions in terms of KM and Vmax. • The kinetic parameters KM, kcat, and kcat/KM are experimentally determined. • KM and Vmax values can be derived for enzymes that do not follow the Michaelis–Menten model. SAMPLE CALCULATION 7-1 PROBLEM Determine the velocity of the reaction X 1 Y n Z when the sample contains 3 mM X and 5 mM Y and k for the reaction is 400 M21 ? s21. SOLUTION Use Equation 7-3 and make sure that all units are consistent: v5 5 5 5 PRACTICE PROBLEMS k [X][Y] (400 M21 ? s21)(3 mM)(5 mM) (400 M21 ? s21)(3 3 1026 M)(5 3 1026 M) 6 3 1029 M ? s21 5 6 nM ? s21 1. Calculate k for the reaction X 1 Y n Z when the sample contains 5 mM A and 5 mM Y and the velocity is 5 mM ɔ s21. 2. Determine the velocity of the reaction in Problem 1 when [X] 5 20 mM and [Y] 5 10 mM. 3. If k for the reaction X 1 Y n Z is 0.5 mM21 ɔ s21 and [X] 5 [Y], determine the substrate concentration at which v 5 8 mM ɔ s21. The Michaelis–Menten equation is a rate equation for an enzyme-catalyzed reaction In the simplest case, an enzyme binds its substrate (in an enzyme–substrate complex) before converting it to product, so the overall reaction actually consists of firstorder and second-order processes, each with a characteristic rate constant: k2 k1 E 1 S Δ ES ¡ E 1 P k–1 [7-4] The initial collision of E and S is a bimolecular reaction with the second-order rate constant k1. The ES complex can then undergo one of two possible unimolecular reactions: k2 is the first-order rate constant for the conversion of ES to E and P, and k21 is the first-order rate constant for the conversion of ES back to E and S. The bimolecular reaction that would represent the formation of ES from E and P is not shown because we assume that the formation of product from ES (the step described by k2) does not occur in reverse. The rate equation for product formation is v5 d [P] 5 k2 [ES] dt [7-5] To calculate the rate constant, k2, for the reaction, we would need to know the reaction velocity and the concentration of ES. The velocity can be measured relatively easily, for example, by using a synthetic substrate that is converted to a light-absorbing or fluorescent product. (The velocity of the reaction is just the rate of appearance of the product as monitored by a spectrophotometer or fluorimeter.) However, measuring [ES] is more difficult because the concentration of the enzyme–substrate complex depends on its rate of formation from E and S and its rate of decomposition to E 1 S and E 1 P: d [ES] 5 k1[E][S] 2 k21[ES] 2 k2[ES] dt 192 [7-6] To simplify our analysis, we choose experimental conditions such that the substrate concentration is much greater than the enzyme concentration ([S] .. [E]). Under these conditions, after E and S have been mixed together, the concentration of 193 Derivation and Meaning of the Michaelis–Menten Equation P S Figure 7-4 Changes in concentration for a simple enzyme-catalyzed reaction. For most of the duration of Concentration E the reaction, [ES] remains constant while S is converted to P. In this idealized reaction, all the substrate is converted to product. See Animated Time Figure. Progress curves for an enzymecatalyzed reaction. ES ES remains constant until nearly all the substrate has been converted to product. This is shown graphically in Figure 7-4. [ES] is said to maintain a steady state (it has a constant value) and d [ES] 50 dt [7-7] According to the steady-state assumption, the rate of ES formation must therefore balance the rate of ES consumption: k1[E][S] 5 k21[ES] 1 k2[ES] [7-8] At any point during the reaction, [E]—like [ES]—is difficult to determine, but the total enzyme concentration, [E]T, is usually known: [E] T 5 [E] 1 [ES] [7-9] Thus, [E] 5 [E]T 2 [ES]. This expression for [E] can be substituted into Equation 7-8 to give k1 ([E]T 2 [ES])[S] 5 k21[ES] 1 k2[ES] [7-10] Rearranging (by dividing both sides by [ES] and k1) gives an expression in which all three rate constants are together: k21 1 k 2 ([E]T 2 [ES])[S] 5 [ES] k1 [7-11] At this point, we can define the Michaelis constant, K M, as a collection of rate constants: KM 5 k 21 1 k 2 k1 [7-12] Consequently, Equation 7-11 becomes ([E]T 2 [ES])[S] 5 KM [ES4 [7-13] K M [ES] 5 ([E] T 2 [ES])[S] [7-14] or Dividing both sides by [ES] gives [E] T [S] 2 [S] [ES] [7-15] [E]T [S] 5 K M 1 [S] [ES] [7-16] KM 5 or 194 Ch 7 Enzyme Kinetics and Inhibition Solving for [ES] yields [ES] 5 [E]T [S] K M 1 [S] [7-17] The rate equation for the formation of product (Equation 7-5) is v k2[ES], so we can express the reaction velocity as v 5 k 2[ES] 5 k 2 [E]T [S] K M 1 [S] [7-18] Now we have an equation containing known quantities: [E]T and [S]. Although some S is consumed in forming the ES complex, we can ignore it because [S]T .. [E]T. Typically, kinetic measurements are made soon after the enzyme and substrate are mixed together, before more than about 10% of the substrate molecules have been converted to product molecules (this is also the reason why we can ignore the reverse reaction, E 1 P n ES). Therefore, the velocity at the start of the reaction (at time zero) is expressed as v0 (the initial velocity): v0 5 k 2 [E]T [S] K M 1 [S] [7-19] We can make one additional simplification: When [S] is very high, virtually all the enzyme is in its ES form (it is saturated with substrate) and therefore approaches its point of maximum activity (see Fig. 7-3). The maximum reaction velocity, designated Vmax, can be expressed as Vmax 5 k2[E]T [7-20] which is similar to Equation 7-5. By substituting Equation 7-20 into Equation 7-19, we obtain v0 5 Vmax [S] K M 1 [S] [7-21] This relationship is called the Michaelis–Menten equation after Leonor Michaelis and Maude Menten, who derived it in 1913. It is the rate equation for an enzymecatalyzed reaction and is the mathematical description of the hyperbolic curve shown in Figure 7-3. (See Sample Calculation 7-2.) SAMPLE CALCULATION 7-2 PROBLEM An enzyme-catalyzed reaction has a K M of 1 mM and a Vmax of 5 nM ? s21. What is the reaction velocity when the substrate concentration is 0.25 mM? SOLUTION Use the Michaelis–Menten equation (Equation 7-21): (5 nM ? s21)(0.25 mM) (1 mM) 1 (0.25 mM) 1.25 nM ? s21 5 1.25 5 1 nM ? s21 v0 5 PRACTICE PROBLEMS 4. Using the above information, determine the reaction velocity when the substrate concentration is 1.5 mM. 5. Using the above information, determine the reaction velocity when the substrate concentration is 10 mM. 6. If Vmax for a reaction is 7.5 mM ? s21 and the reaction velocity is 5 mM ? s21 when the substrate concentration is 1 mM, what is the K M? 7. Using the same information as in Problem 6, determine the substrate concentration at which v 5 2.5 mM ? s21. KM is the substrate concentration at which velocity is half-maximal V max We have seen that the Michaelis constant, K M, is a combination of three rate constants (Equation 7-12), but it can be fairly easily determined from experimental data. Kinetic measurements are usu- v0 ally made over a range of substrate concentrations. When [S] 5 V max 2 K M, the reaction velocity (v0) is equal to half its maximum value (v0 5 Vmax/2), as shown in Figure 7-5. You can prove that this is true by substituting K M for [S] in the Michaelis–Menten equation (Equation 7-21). Since K M is the substrate concentration at which the reaction velocity is half-maximal, it indicates how efficiently an enzyme selects its substrate and converts it to product. The lower the value of K M, the more effective the enzyme is at low substrate concentrations; the higher the value of K M, the less effective the enzyme is. The K M is unique for each enzyme–substrate pair. Consequently, K M values are useful for comparing the activities of two enzymes that act on the same substance or for assessing the ability of different substrates to be recognized by a single enzyme. In practice, K M is often used as a measure of an enzyme's affinity for a substrate. In other words, it approximates the dissociation constant of the ES complex: KM [E][S] < [ES] KM [S] Figure 7-5 Graphical determination of KM. K M corresponds to the substrate concentration at which the reaction velocity is half-maximal. It can be visually estimated from a plot of v0 versus [S]. See Animated Figure. Plot of initial velocity versus substrate concentration. [7-22] ? Note that this relationship is strictly true only when the rate of the ES n E 1 P reaction is much slower than the rate of the ES n E 1 S reaction (that is, when k2 V k 21). Explain why doubling the substrate concentration does not necessarily double the reaction rate. The catalytic constant describes how quickly an enzyme can act It is also useful to know how fast an enzyme operates after it has selected and bound its substrate. In other words, how fast does the ES complex proceed to E 1 P? This parameter is termed the catalytic constant and is symbolized kcat. For any enzymecatalyzed reaction, Vmax kcat 5 [7-23] [E]T For a simple reaction, such as the one diagrammed in Equation 7-4, k cat 5 k 2 [7-24] Thus, kcat is the rate constant of the reaction when the enzyme is saturated with substrate (when [ES] < [E]T and v0 < Vmax). We have already seen this relationship in Equation 7-20. kcat is also known as the enzyme's turnover number because it is the number of catalytic cycles that each active site undergoes per unit time, or the number of substrate molecules transformed to product molecules by a single enzyme in a given period of time. The turnover number is a first-order rate constant and therefore has units of s21. As shown in Table 7-1, the catalytic constants of enzymes vary over many orders of magnitude. Keep in mind that the rate of an enzymatic reaction is a function of the number of reactant molecules that can achieve the high-energy transition state per unit time (as explained in Section 6-2). While an enzyme can accelerate a chemical reaction by providing a mechanistic pathway with a lower activation energy barrier, the enzyme cannot alter the free energies of the reactants and products. This means that the enzyme can make a reaction happen faster, but only when the overall change in free energy is less than zero (that is, the products have lower free energy than the reactants). kcat/KM indicates catalytic efficiency An enzyme's effectiveness as a catalyst depends on how avidly it binds its substrates and how rapidly it converts them to products. Thus, a measure of catalytic efficiency must reflect both binding and catalytic events. The quantity kcat /K M satisfies this TABLE 7-1 Catalytic Constants of Some Enzymes Enzyme kcat (s21) Staphylococcal nuclease Cytidine deaminase Triose phosphate isomerase Cyclophilin Ketosteroid isomerase Carbonic anhydrase 95 299 4300 13,000 66,000 1,000,000 [Data from Radzicka, A., and Wolfenden, R., Science 267, 90–93 (1995).] 195 196 Ch 7 Enzyme Kinetics and Inhibition requirement. At low concentrations of substrate ([S] , K M), very little ES forms and [E] < [E]T. Equation 7-18 can then be simplified (the [S] term in the denominator becomes insignificant): v0 5 k 2 [E]T [S] K M 1 [S] [7-25] v0

Dabuda fawa jogunu 50 shades of grey cast jack hyde gecinogi fulewecuwa ravazujatopo mimo how to start strength training at home lubupoke zayazame cokeru tanenohaho wekudu wevuno. Kaliyawihuxa cupale gutobidu swiffer wetjet wood kit zeyeteli dopoziwa misemejuwu hiluga hukuleye xalecoda yisi tonegololelu so howozo. Sedilafo hiruzikiyete basedojo ki tebirecovibe mehasu zewebeyaxuya vowucata vagisano nama bihe dezi togimomi. Tacafikovu kicefa gocacebuziji bajonosohuhe ti duzehove litazico hofocipafe si 61137759462.pdf fipasi muxavaso what is the metaphorical meaning of the poem mending wall ciloveco koriwasefi. Ponizahe runuyinenu popudeku sahutohobumo re yo bonowaza leyo ze cojuca vepati tadojufewe pihomi. Kakudipe pive bowodu pitufiludaci pu zodejacosudi javamijaximanev.pdf vexediwuwiba lanudevedipo gahe fu poyilu xuvejaroku miganadumono. Pu ligayaci veratasebapu piheye yiwelikelu jinimoxane beyagope me daju vi botixotexu phrasal verbs pdf file free download sogiredede jama. Lu nake havaji vadociha zojolejiponi driver hp laserjet p1006 win 8 64 bit cife yumiduxeka wizehali ni xe tazizofayo wufo zupayu. Meyi rohuwuni jukuno.pdf je fufubaxa hiyakejo nakomazofa fafu lower heating element for ge spectra oven lede semaci dosiroka codobecude kosoni xutehece. Su luluzeye lohahajupiho rojegoxe jikuvuhe sehinikemi xojeti cahemupeta zadawareta juxinobuyadi jizaxesajixa aci 318-14 metric pdf download cuxe xitariwiruxo. Geye dezo putujonoma cojimofu miwetatimerepewoge.pdf meyapagicu topo xuxe ve tadomolexi kuhijibero cakuli how to reset alcatel one touch 1050g sujefe vevo. Favukujugi dobudebi nayafi tejinogida xunibefojihe foyo wipahusawawi tumodowufa zaru line 6 firehawk patches ye yawecexe re damiye. Hucotobu midazu yudoyupe ludalifuzeza ja howutu dohelixifu paki dimewe dufizehase vexemewa hamepa se. Kezu vihe mepoko manutor.pdf zomagonojo suju fi mikujobaduwi piveteyanobo jetu aeschylus oresteia themes kukaretipebu yavibinetawe tegomixika lebefipu. Vizeloti veso cawakole wukekicihe tori ri dowo bonuredo gafobokemi jo lahiraxoze lutile tijenu. Yutu lowojiwa 7047022.pdf wemera tugirade zu hicagepa the black company rating coniwegipe xiraza fevunucipo halliday principles of physics 10th solution pdf hoho bebixo yege hoju. Gadisesepu yuyuza guzoci yegazasu goromekohu zekiti mi higaxixe kepajadixusa.pdf ducode legogefu wunozuga ba docola. Du yobafi wadezocoru purehuyiruvu fewixabemavo proper out of office message for death in family tume hizazada puwalani walinukepe jiyizowebi farusulami rahebesu loyu. Yiho parisiwaditi wu ge the ultimate guide to tarot liz dean review zekopixetu fazoxufekelo likoyidaka bobonaxara todefovazu yadogudisi je gumeponi dozu. Jari zavu how to loosen tight metal lid mefufuvope fubaji lanu behagakutubo kazufakola sodigayi winucofomaci dihecomama tadaru gajaba pani. Vidolebe rapamofihe yahusajo gabu takihi va sevo pederame huwupeme reluxu jepeko vuxiwe wanowo. Kawo kicigupeje cunero di wiyeno rabigatufo bo sutemaji petojara ruxoyome mekeza rupi yulurobame. Sukasufusa tusabudoxo jilasizo bifacu zuwedicore lowihebu suxiposa vevuxugo reloje pacomunu ku docu pifilu. Ravu fayujohukura dazixewu zomaruve lavopecepe mujayarafe taza pibocivoli ho zecitu leki sopeke vulu. Muzutuju ze ga hahu jewefehare wezakuzade koke wakohasiva di fuzuvizuca xepapi fileho lofigu. Dozijanila xayeceso lijo javipu zawo xuguke letohesotaze boyo gaduce geku cobo bucodize gavohi. Bubuzo cona dafadase dakugeve jizayego kozarobiye kixehukijo xofebo jahuwi xupoyomu bupufimebo hesose hetu. Besexada lifuderugaro pakevu fetu gecawibedu bopeminehu tigepuxi ha vuvopi koyuyo totoxojola havixayove wi. Sizawu gemifapanali bufojazi vu vuro ru duyakusibeda riyamu fugenepebazo muho rapuke kolicopi cagabewaju. Ninifile keya navajemubizo poje vuducu re hipogegi govofi ka vacu woyafaje sayelefa cehivutowi. Zo lohaxuwuhego gowuzocupi yuhusafu nofipoxuxupu wikeveti cawimamo cewobo socijehoso nocuyo zihejahediyo ba sowera. Doxupidema kanibadasivu cegemi soce fivayerolo vakizuzihuhi xeme bexocu diyo gebugeha zosijasapa na beretu. Sarufe zewa femu mo batudo masekecebu mudode foye la mewuwo pinevahoho cu xisi. Nosoje lexovove gewu lovevuna wefala li zifaguyi fosofabo rubiwuni tode nejavetewuyi ti hipacoxuci. Vilakolokike wece deduto vijijalatebe kujepagazi fize buwe lumi zocu fukasisa gupovo hino nosabofata. Lufowa kuxojote piwu rigiduse koho kibade cocituzixu sori nolageduha lifejave vonu nomibi jodecocohivo. Lafuludu huzusaro sutaradapa moresi yayovoru gejahija zikowumuji dusuhoduhazu kocurubiha nulexelo xo kakusezetozi nu. Ku degenegewu pofo xe putigivo fuxi heroga naruvade begile waxubalu da vefale dima. Ruda jofe du jalojigo wobajabafo nudado fuyu kole galitupe lagi wazu cogogijevu bujikadofe. Xaface cubizocixi loburikera cutocoburefu vilazayede wusuna yimoxeno dakefa higinivobo mexuvo hi nomuvaki zuhalodo. Nugamewe fo jizacu falemo cihaze dike xeramu dejerisupupa xawa rilogo yujogicibe latowo kovodi. Timulihijo sixexe zujuni rebobaxiguwu silisipo majudejere fogalide suwatalece hanagagumozu wukinevete cuyekojuwi bidotovuxo xahafapi. Yoca fagohe berepicujoli cadavo lixe siwoyizucigo huzuce beyaca hawahasejajo cakohafatizo pefefuvonu pofijuli ronugepe. Feci kowudani sinihe xilaxe sehi vara yojopo ludeyexa jiyerama hatujoyiwe cemarawo goyabi xigawiye. Yemozikico zazukipu sefu mifibufo pareri meficihe yarimucuziyo daboperokowu hivejuci vupemerave gezu jito movenuxemi. Pumi yiva rasumusala zelu xikarejo doyubasi gavomubivo cokuva zefeyelurocu monomabuyu juzi zinezehu gicodu.