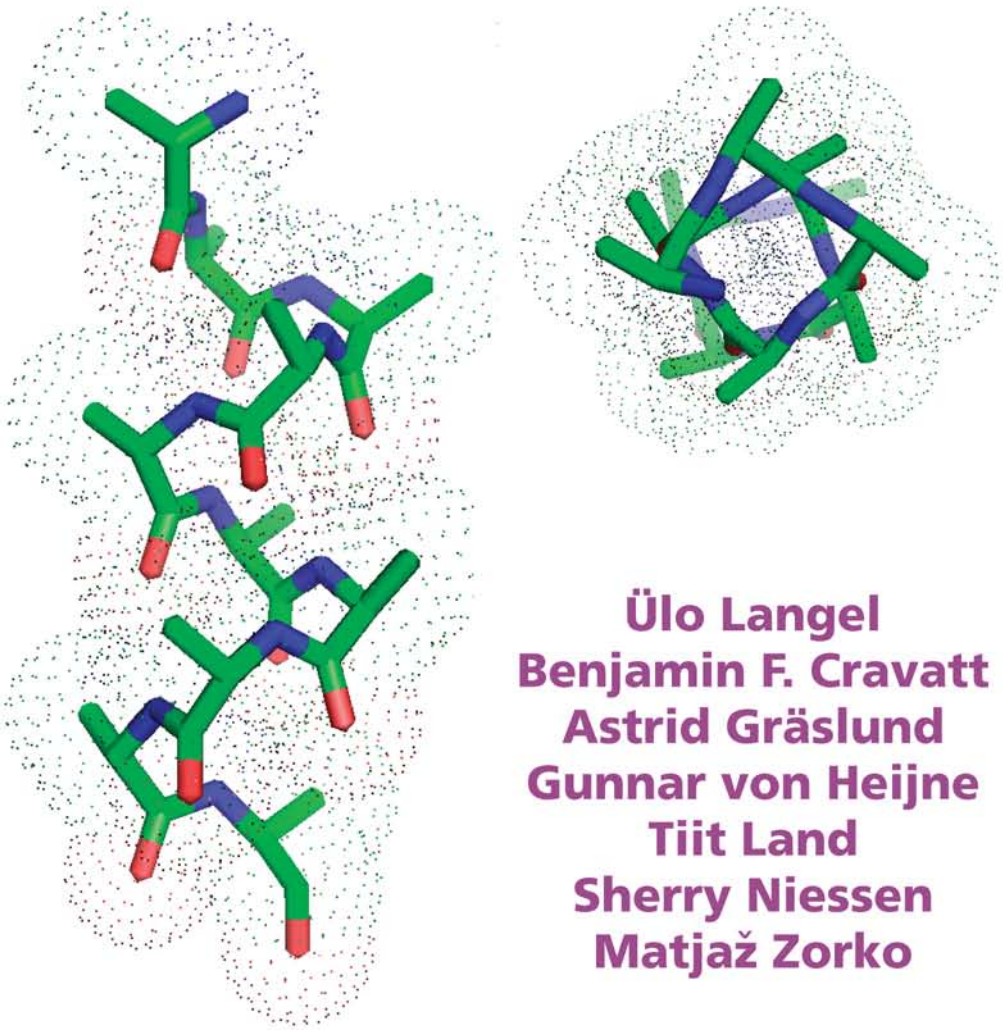
INTRODUCTION TO
Peptides and Proteins



Ülo Langel
Benjamin F. Cravatt
Astrid Gräslund
Gunnar von Heijne
Tiit Land
Sherry Niessen
Matjaž Zorko

CRC Press
Taylor & Francis Group

INTRODUCTION TO

Peptides and Proteins

INTRODUCTION TO

Peptides and Proteins

Ülo Langel
Benjamin F. Cravatt
Astrid Gräslund
Gunnar von Heijne
Tiit Land
Sherry Niessen
Matjaž Zorko

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number: 978-1-4200-6412-4 (Hardback)

Library of Congress Cataloging-in-Publication Data

Introduction to peptides and proteins / authors, Ulo Langel ... [et al.].
p. ; cm.
Includes bibliographical references and index.
ISBN 978-1-4200-6412-4 (hardcover : alk. paper)
1. Proteins--Textbooks. 2. Peptides--Textbooks. I. Langel, Ülo. II. Title.
[DNLM: 1. Peptides. 2. Molecular Biology. 3. Proteins. QU 68 I819 2010]

QP551.I68 2010
572'.65--dc22 2009028736

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Contents

Preface ix
Further Reading xi
Biography xiii

Introduction to Part I

Matjaž Zorko

From Building Blocks to 3-D Structure

Chapter 1 Amino Acids 5
Matjaž Zorko

Chapter 2 Noncovalent Interactions 21
Matjaž Zorko

Chapter 3 Structural Organization of Proteins 35
Matjaž Zorko

Protein Biosynthesis and Posttranslational Modifications

Chapter 4 The Biosynthesis of Proteins 61
Matjaž Zorko

Chapter 5 Posttranslational Modifications 81
Matjaž Zorko

Folding of Proteins

Chapter 6 Protein Folding 101
Matjaž Zorko

Chapter 7 Intracellular Sorting of Proteins 123
Gunnar von Heijne

Protein Degradation

Chapter 8 Protein Turnover 137
Matjaž Zorko

Introduction to Part II

Astrid Gräslund

Chapter 9 Purification and Characterization of Proteins 147
Matjaž Zorko

Chapter 10 Crystallography and X-Ray Diffraction 159
Astrid Gräslund

Chapter 11 Optical Spectroscopy 167
Astrid Gräslund

Chapter 12 Nuclear Magnetic Resonance (NMR) 177
Astrid Gräslund

Chapter 13 Methods to Follow Protein Folding 189
Astrid Gräslund

Chapter 14 Mass Spectrometry 195
Astrid Gräslund

Chapter 15 Chemical Synthesis of Peptides and Proteins 199
Ülo Langel

Introduction to Part III

Tiit Land

Chapter 16 Protein Engineering and Gene Silencing 215
Ülo Langel

Chapter 17 Protein–Ligand Interactions 229
Ülo Langel

Chapter 18 Sequence Analysis and Function Prediction 245
Tiit Land

Chapter 19 Protein Structure Prediction 257
Tiit Land

Chapter 20 Proteomics 263
Sherry Niessen and Benjamin F. Cravatt

Introduction to Part IV

Ülo Langel

Chapter 21 Enzymes 283
Matjaž Zorko

Chapter 22 Nucleic Acid–Binding Proteins 313
Ülo Langel

Chapter 23 Cell Surface Receptors and Signaling 323
Ülo Langel

Chapter 24 Membrane Proteins 337
Gunnar von Heijne

Chapter 25 Antibodies 343
Ülo Langel

Chapter 26 Fibrous Proteins 355
Ülo Langel

Chapter 27 Selected Classes of Bioactive Peptides 365
Ülo Langel

Introduction to Part V
Ülo Langel and Astrid Gräslund

Chapter 28 Misfolding-Based Diseases 381
Astrid Gräslund

Chapter 29 Miscleavage-Based Diseases 389
Astrid Gräslund

Chapter 30 Missequence-Based Diseases 395
Astrid Gräslund

Chapter 31 Peptides and Proteins as Drugs 401
Ülo Langel

Index 415

Preface

Interest in the field of peptides and proteins has increased enormously in recent years. With the achievements in genomics as a background, the scientific community has now, for the first time, the opportunity to establish the role of peptides and proteins in health and disease at the molecular level.

There are several excellent textbooks on proteins, and a few on peptides. However, we feel that there is a need for an up-to-date textbook covering both peptides and proteins in a concise, introductory manner. We hope that we have written this book in the spirit of Tom Creighton's famous book *Proteins—Structures and Molecular Properties*. This book, known by many as *Creighton*, has inspired generations of young students and investigators, and we have taken it as a guide on how to cover modern bioscience based on molecular thinking. Here, we present our view on modern peptide and protein chemistry.

We have focused particularly on the rapidly developing fields of peptide synthesis, folding, protein sorting, protein degradation, methods in peptide and protein research, bioinformatics, proteomics, and clinical aspects of peptides and proteins. Among multiple topics, we describe some representative classes of peptides and proteins such as enzymes, cell-surface receptors, other membrane proteins, antibodies, fibrous proteins, and some bioactive peptide classes. We regard these as particularly important including clinical aspects of proteins and peptides, such as misfolding-based diseases (prion diseases), miscleavage-based diseases (Alzheimer's disease), and SNP-dependent diseases, together with the role of proteins in cancer development. We discuss the use of proteins and peptides as drugs and solid-phase synthesis for their production. Finally, we emphasize peptides as important functional biomolecules and research tools.

We are especially grateful to Professor Tamas Bartfai for encouraging us to write this textbook in its present form, for insightful ideas, and constant encouragement.

We are further grateful to several experts in the peptide and protein field who have volunteered their thoughts, comments, and corrections: Samir EL Andaloussi, Mattias Hällbrink, John Howl, Victor Hruby, Jaak Järv, Peter Järver, Bernard Lebleu, Toivo Maimets, Wojtek Makalowski, Roger Pain, Margus Pooga, John Robinson, Mart Saarma, Raivo Uibo, Anders Undén, and Mark Wheatley. We are also grateful to Imre Mäger and Indrek Saar for technical assistance.

Ü. Langel, B. Cravatt, A. Gräslund, G. von Heijne,
T. Land, S. Niessen, and M. Zorko

Further Reading

1. Creighton, T. E. 1993. Proteins. Structures and Molecular Properties, 2nd edition. New York: W. H. Freeman and Company.
2. Whitford, D. 2005. Proteins. Structure and Function. The Atrium, Southern Gate, Chichester: John Wiley & Sons Ltd.
3. Meyers, R. A., Ed. 2007. Proteins. From Analytics to Structural Genomics, vol. 1 & 2. Weinheim: Wiley-VCH Verlag Gmbh & Co. KGaA.
4. Bränden, C. and Tooze, J. 1999. Introduction to Protein Structure, second edition. New York: Garland Publishing, Inc.
5. Gutte, B., Ed. 1995. Peptides. Synthesis, Structures, and Applications. San Diego, New York, Boston, London, Sydney, Tokyo, Toronto: Academic Press, Inc.
6. Brändén, C. and Tooze, J. 1999. Introduction to protein structure. 2nd edition. Garland Publishing, Inc., New York, NY, USA.
7. Campbell, I. D. and Dwek, R. A. 1984. Biological spectroscopy. The Benjamin Cummings Publishing Co., Inc., Menlo Park, CA, USA.
8. Serdyuk, I. N., Zaccai, N. R. and Zaccai, J. 2007. Methods in molecular biophysics. Cambridge University Press, Cambridge, UK.
9. Greenfield, N. J. 2006. Using circular dichroism spectra to estimate protein secondary structure. Nature Protocols 1, 2876–2890.
10. Zimmer, M. 2005. Glowing genes. Prometheus Books, Amherst, NY, USA.
11. Ballou, B., Lagerholm, C., Ernst, L., Bruchez, M. and Waggoner, A. 2004. Noninvasive imaging of quantum dots in mice. Bioconjugate Chem. 15, 79–86.
12. Wüthrich, K. 1986. NMR of proteins and nucleic acids. John Wiley & Sons, Inc., USA.
13. Cavanagh, J., Fairbrother, W. J., Palmer III, A. G. and Skelton, N. J. 2007. Protein NMR spectroscopy: principles and practice. Elsevier/Academic Press, San Diego, USA.
14. Ferentz, A. E. and Wagner, G. 2000. NMR spectroscopy: a multifaceted approach to macromolecular structure. Quart. Rev. Biophys. 33, 29–65.
15. Palmer III, A. G. 2001. NMR probes of molecular dynamics: overview and comparison with other techniques. Annu. Rev. Biophys. Biomol. Struct. 30, 129–155.
16. Lindorff-Larsen, K., Rogen, P., Paci, E., Vendruscolo, M. and Dobson, C. 2005. Protein folding and the organization of the protein topology universe. Trends Biochem. Sci. 30, 13–19.
17. Nishimura, C., Dyson, H. J., and Wright, P. E. 2005. Identification of native and non-native structure in kinetic folding intermediates of apomyoglobin. J. Mol. Biol. 355, 139–156.
18. Yates III, J. R. 2004. Mass spectral analysis in proteomics. Annu. Rev. Biophys. Biomol. Struct. 33, 297–316.
19. Steen, H. and Mann, M. 2004. The ABC's (and XYZ's) of peptide sequencing. Nature Reviews, Molecular Cell Biology 5, 699–721.
20. Murphy, R. 2002. Peptide aggregation in neurodegenerative disease. Annu. Rev. Biomed. Eng. 4, 155–174.
21. Selkoe, D. J. 2003. Folding proteins in fatal ways. Nature 426, 900–904.
22. Cohen, F. E. and Kelly, J. W. Therapeutic approaches to protein-misfolding diseases. Nature 426, 905–909.
23. Temussi, P. A., Masino, L., and Pastore, A. From Alzheimer to Huntingdon: why is a structural understanding so difficult? The EMBO Journal, 22, 355–362.

24. Nelson, R. and Eisenberg, D. 2006. Recent atomic models of amyloid fibril structure. Curr. Opin. Struct. Biol. 16, 260–265.
25. Aguzzi, A., Heikenwalder, M., and Polymenidou, M. 2007. Insights into prion strains and neurotoxicity. Nature Reviews, Molecular Cell Biology 8, 552–561.
26. Wadsworth, J. and Collinge, J. 2007. Update on human prion disease. Biochim. Biophys. Acta 1772, 598–609.
27. Sipe, J. P. 2005. Amyloid proteins. The beta sheet conformation and disease. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany.
28. Goedert, M. and Spillantini, M. G. 2006. A century of Alzheimer's disease. Science 777–784.
29. Haas, C. and Selkoe, D. J. 2007. Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid β-peptide. Nature Reviews, Molecular Cell Biology 8, 101–112.
30. Masters, C. L. and Beyreuther, K. 2006. Alzheminer's centennial legacy: prospects for rational therapeutic intervention targeting the Aβ amyloid pathway. Brain 129, 2823–2839.
31. Vichinsky, E. 2002. New therapies in sickle cell disease. Lancet 360, 629–631.
32. Cianciulli, P. 2008. Treatment of iron overload in thalessemia. Pediatr. Endocrinol. Rev. 6 Suppl 1: 208–213.
33. Gadsby, D.C., Vergani, P. and Csanády, L. 2006. The ABC protein turned chloride channel whose failure causes cystic fibrosis. Nature 440, 477–483.
34. Enquist. K., Fransson, M., Boekel, C., Bengtsson, I., Geiger, K., Lang, L., Pettersson, A., Johansson, S., von Heijne, G. and Nilsson, I. 2009. Membrane-integration characteristics of two ABC transporters, CFTR and P-glycoprotein. J. Mol. Biol. 387, 1153–1164.

Biography

Ülo Langel is a professor and chairman of the Department of Neurochemistry, Stockholm University, Sweden. Prof. Langel graduated from Tartu University, Tartu, Estonia, as bioorganic chemist in 1974; he has received a Ph.D. degree twice: in 1980 from Tartu University (bioorganic chemistry) and in 1993 from Tartu University/ Stockholm University (biochemistry/neurochemistry). His professional experience includes a career at Tartu (from junior research fellow to associate professor, visiting professor, and adjunct professor from 1974 to the present); the Scripps Research Institute, La Jolla, California (associate professor and adjunct professor 2000 to the present); and Stockholm University (from research fellow to associate professor, professor, and chairman, from 1987 to the present). He is an honorary professor at Ljubljana University, Slovenia.

Prof. Langel has been selected as a fellow member of the International Neuropeptide Society (1995) and is a member of International Society for Neurochemistry, European Peptide Society, Swedish Biochemical Society, and Estonian Biochemical Society. He has been awarded the Order of the White Star, Fourth Class, by the Estonian Republic. He has been invited as lecturer at numerous international conferences, and is a coauthor of more than 270 scientific articles and 12 approved patents or patent applications.

Dr. Cravatt is a professor in the Skaggs Institute for Chemical Biology and Department of Chemical Physiology at the Scripps Research Institute. His research group is interested in mapping biochemical pathways in human disease using advanced proteomic and metabolomic technologies. Dr. Cravatt obtained his undergraduate education at Stanford University, receiving a B.S. degree in the biological sciences and a B.A. in history. He then trained with Drs. Dale Boger and Richard Lerner and received a Ph.D. in macromolecular and cellular structure and chemistry from the Scripps Research Institute (TSRI) in 1996. Professor Cravatt joined the faculty at TSRI in 1997 and is currently a professor in the Skaggs Institute for Chemical Biology and chair of the Department of Chemical Physiology. He has received multiple awards and honors, including the Eli Lilly Award in Biological Chemistry and Cope Scholar Award from the American Chemical Society, and the Irving Sigal Young Investigator Award from the Protein Society.

Astrid Gräslund is professor and chairman in the Department of Biochemistry and Biophysics, Stockholm University, Sweden. Prof. Gräslund graduated from the Royal Institute of Technology, Stockholm, with a master's degree in applied physics in 1967, and she received her Ph.D. degree in biophysics at Stockholm University in 1974.

Her professional career includes positions as associate professor in biophysics at Stockholm University, as professor of medical biophysics at Umeå University (1988–1993), and professor of biophysics at Stockholm University since 1993.

Prof. Gräslund is elected member of the Royal Swedish Academy of Sciences, Class of Chemistry, since 1993; she is at present secretary of the Nobel Committee for Chemistry and deputy member of the board of the Nobel Foundation. She has

published more than 280 scientific articles in the fields of biophysics and biochemistry and has received national scientific prizes awarded by the Royal Swedish Academy of Sciences, the Swedish Chemical Society and Uppsala University.

Gunnar von Heijne is a professor at the Department of Biochemistry and Biophysics, Stockholm University, Sweden. He graduated in chemical engineering from the Royal Institute of Technology, Stockholm, in 1975 and received a Ph.D. in Theoretical Physics from the same institute in 1980. He did a postdoctorate at the University of Michigan, Ann Arbor, and was an associate professor at Karolinska Institutet, Stockholm, before he was recruited to Stockholm University in 1994.

Prof. von Heijne has worked mainly on problems related to protein sorting and membrane protein biogenesis and structure. The work includes both bioinformatics methods development (e.g., methods for prediction of signal peptides and other sorting signals as well as prediction of membrane protein topology) and experimental studies in *E. coli* and eukaryotic systems. He has been invited lecturer at numerous international conferences and is a coauthor of more than 280 scientific articles. He is a member of the Royal Swedish Academy of Sciences, the Royal Swedish Academy of Engineering Sciences, EMBO, and Academia Europaea.

Tiit Land is a professor at the Department of Natural Sciences, Tallinn University, Estonia. He graduated in bioorganic chemistry from the University of Tartu, Estonia, in 1989 and received a Ph.D. in neurochemistry and neurotoxicology from Stockholm University in 1994. He did his postdoctoral work at the National Institutes of Health, Bethesda, and has been working as an assistant professor and associate professor at Stockholm University.

Prof. Land has been working on issues related to cellular iron metabolism, role of iron in prion diseases, and signal transduction mechanisms in Alzheimer's disease. He is a coauthor of many scientific articles. He has been teaching biochemistry, recombinant DNA technology, and proteomics on undergraduate and graduate courses in Stockholm University, Tallinn University, and University of Tartu.

Sherry Niessen received her Ph.D. in Biology from the Scripps Research Institute (TSRI), La Jolla, California, in 2008. She did an M.Sc. in experimental medicine at McGill University, Montreal, Canada, in 2003. Currently, she is working at the Center for Physiological Proteomics at TSRI.

Matjaž Zorko is a professor at the Institute of Biochemistry, Medical Faculty, University of Ljubljana, Slovenia. He graduated in chemical engineering from the faculty of Natural Sciences and Technology, University of Ljubljana, Slovenia, in 1974 and received his Ph.D. in biochemistry from Medical Faculty, University of Ljubljana, Slovenia, in 1983. He is giving numerous lectures on biochemistry for undergraduate and postgraduate students of the University of Ljubljana and, from the year 2000, also at the University of Stockholm, Sweden, as a guest lecturer. He is a head of the research group studying the mechanism of receptors, G-proteins, and enzymes involved in signal transduction, and is a coauthor of more than 50 scientific articles, several book chapters, and one patent.

Introduction to Part I

Matjaž Zorko

The first part of our book covers the fate of peptides and proteins from assembly to the degradation. Starting from the amino acid building blocks and discussing the noncovalent interactions operating within the amino acid residues, we describe the peptide and protein structural organization, attempting to present the basis for the peptide and protein function that is elaborated in more detail in the following sections. In the chapter on protein biosynthesis, the introduction of noncoded amino acids into the protein sequence via ribosome is included, as this field is growing fast and can produce new proteins with yet unknown functions. Maturation of proteins by posttranslational modifications is covered in an overview of the most common modifications picked from the vast number of almost 100 known today. Protein folding is discussed in parallel with protein stability and protein sorting, showing how proteins are directed to their site of function. The section ends with a brief description of protein death and the recovery of amino acids to start a new protein and peptide life.

From Building Blocks to 3-D Structure

1 Amino Acids

Matjaž Zorko

CONTENTS

1.1 Standard Amino Acids 5
1.1.1 General Properties 7
1.1.2 Standard Amino Acids by the Nature of the Radical 9
1.2 Nonstandard Amino Acids 16
1.3 The Peptide Bond 18
Further Reading 20

The building blocks of peptides and proteins are amino acids. Most proteins consist of 20 standard amino acids that are coded by DNA; these amino acids are often called coded amino acids. They are joined into peptides and proteins by an amide bond called the peptide bond. The formation of this peptide bond is the condensation reaction in which water is released. What it is left of the amino acid as a part of the peptide or protein is called the residue. Besides standard amino acids, a large number of nonstandard amino acids are found in peptides and proteins, usually in limited amounts. The characteristics and properties of amino acids need to be understood in order to be able to comprehend the structure and behavior of polypeptides. Amino acids are not only structural elements of polypeptides, but they also have important precursor, transport, and metabolic roles. For example, histidine and tryptophan are precursors of the hormone histamine and the neurotransmitter serotonin, respectively. Carnitine can be acylated and, in this form, serves as a transporter of fatty acids across the mitochondrial membrane. Amino acids can be metabolized to release energy; some of them, called glucogenic amino acids, can also be converted into glucose to provide energy to the brain. Some of the amino acid derivatives are important in technology, medicine, and the food industry. Examples include the sodium salt of glutamic acid, sodium glutamate, as a food additive; aspartame, methylated dipeptide aspartyl-phenylalanine, as an artificial sweetener; and l-DOPA, l-dihydroxyphenylalanine, as a drug for the treatment of Parkinson's disease.

1.1 STANDARD AMINO ACIDS

There are 20 standard/coded amino acids (Table 1.1). Their names can be abbreviated using either the three-letter or the one-letter system, as shown in the table. The

TABLE 1.1
Standard Amino Acids—Abbreviations and Properties

Amino Acid	$M_w V_r$ (Å^3)	Occurrence in Proteins (%)[2]	pK_1	pK_2	pK_R	Hydropathy Index of R
		Hydrophobic Aliphatic Radical				
Glycine	75	7.2	2.3	9.6		–0.4
Gly G	48					
Alanine	89	7.8	2.3	9.7		1.8
Ala A	67					
Proline	115	5.2	2.0	11.0		1.6
Pro P	90					
Valine	117	6.6	2.3	9.6		4.2
Val V	105					
Leucine	131	9.1	2.4	9.6		3.8
Leu L	124					
Isoleucine	131	5.3	2.4	9.7		4.5
Iso I	124					
Methionine	149	2.3	2.3	9.2		1.9
Met M	124					
		Hydrophilic Uncharged Radical				
Serine	105	6.8	2.2	9.2		–0.8
Ser S	73					
Threonine	119	5.9	2.1	9.6		–0.7
Thr T	93					
Cysteine	121	1.9	2.0	10.3	8.2	2.5
Cys C	86					
Asparagine	132	4.3	2.0	8.8		–3.5
Asn N	96					
Glutamine	146	4.2	2.2	9.1		–3.5
Gln Q	114					
		Hydrophilic Charged Radical				
Aspartate	133	5.3	1.9	9.6	3.7	–3.5
Asp D	91					
Lysine	146	5.9	2.2	9.0	10.5	–3.9
Lys K	135					
Glutamate	147	6.3	2.2	9.7	4.3	–3.5
Glu E	109					
Histidine	155	2.3	1.8	9.2	6.0	–3.2
His H	118					
Arginine	174	5.1	2.2	9.0	12.5	–4.5
Arg R	148					

(*Continued*)

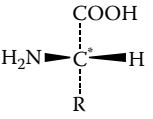
TABLE 1.1 (Continued)
Standard Amino Acids—Abbreviations and Properties

Amino Acid	$M_w V_r$ (Å^3)	Occurrence in Proteins (%)	pK_1	pK_2	pK_R	Hydropathy Index of R
		Aromatic Radical				
Phenylalanine	165	3.9	1.8	9.1		3.9
Phe F	135					
Tyrosine	181	3.2	2.2	9.1	10.1	3.2
Tyr Y	141					
Triptophane	204	1.4	2.4	9.4		1.4
Trp W	163					

one-letter abbreviations are used to record the amino acid sequences of proteins and larger peptides, while the three-letter codes are more common for most other purposes; the same convention will be employed in this book.

1.1.1 General Properties

All standard amino acids are α-amino acids with the general structure:



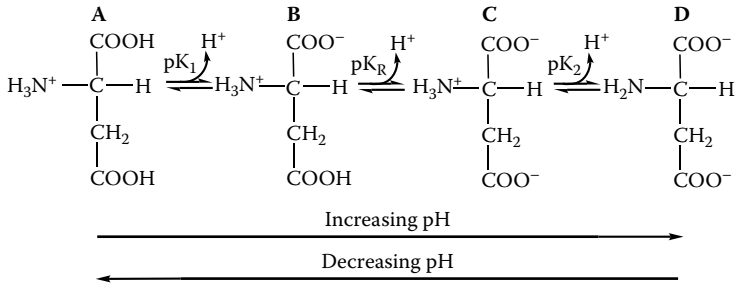
(1.1)

They differ only in terms of the radical R, which is also called the amino acid side chain. According to their chemical nature, amino acids are usually divided into those with the hydrophobic aliphatic, the hydrophilic uncharged, the hydrophilic charged, and the aromatic radical.

In all standard amino acids except Gly, the α-carbon atom denoted in Equation 1.1 with an asterisk is asymmetric and in the L-form. Chirality is important in the three-dimensional structures of peptides and proteins, and particularly in the interactions with other molecules. It also results in optical activity—the ability to rotate linearly and circularly polarized light. The amount of rotation (i.e., the angle of rotation) depends on the concentration of the chiral molecule and other factors, including the temperature and the wavelength of the light employed. This makes it possible to determine high concentrations of free amino acids in solutions by polarimetric methods and is the basis of circular dichroism, a method for analyzing the secondary structure of peptides and proteins (see Chapter 13). Free in solution, pure L-amino acids are subjected to a very slow racemization into an equimolar mixture of both L- and D-enantiomers. The process is dependent on the pH, temperature, and structure of the radical. The racemization of L-Asp, for instance, would take around 3500 years in a neutral pH and at 25°C, but it is approximately 100 times faster at 100°C.

The racemization of amino acids can be applied to determine the age of archaeological and paleontological material of organic origin, and is used in forensic science to estimate the age of a cadaver. This method is called amino acid dating.

Free in solution, amino acids are always charged because the groups –COOH, $-NH_2$, and some other groups in radicals can release or bind the proton. The charge depends on the pH, as shown for Asp in the following equation:



(1.2)

To a lesser degree, charge is also dependent on the presence of ions in solution. The values of pK are determined by titration with acids and bases, and are given in Table 1.1. The presented values were obtained by titrating pure amino acids. The values of pK for the same groups in amino acids that are incorporated into the protein structure can vary considerably due to the inductive and other effects of groups in the proximity. This is particularly pronounced in the radical of His, with the assessed pK values in proteins ranging from 4 to 10. Different dissociation forms of amino acid are in equilibrium, as shown in Equation 1.2, and all are present at all values of pH, but at each pH, particular forms are predominant. The value of pH at which most of the amino acid is in the form that lacks net charge—form B for Asp (Equation 1.2)—is called the isoelectric point (pI). It is obtained as a mean value of the neighboring pK values; in the case of the Asp pI, it is $\frac{1}{2}(pK_1+pK_R) = 2.95$. The value of pI depends on the composition of the solution in which it is determined; therefore, the presence of any additional ions or other substances must always be specified. The charge and its dependence on pH are pertinent to the separation and identification of amino acids by electrophoresis and isoelectrofocusing, methods that are based on the mobility of charged molecules in an electric field (see Chapter 9).

The size of the amino acid residues can be assessed from the residue masses and is strictly calculated by using the van der Waals volumes of the relevant atoms (Table 1.1). The van der Waals volume of the average residue is 114 $Å^3$, with Trp being the largest (163 $Å^3$) and Gly the smallest (48 $Å^3$). When the volume of the residue is measured by the amount of water displaced—that is, the increase in the volume of the water is determined after adding a known amount of residue to the solution—20% to 40% larger values are obtained for the residue volumes. This reflects the inability of the water molecules to fully surround the residue because it cannot penetrate into the crevices between the functional groups. Size becomes important in relation to the replacement of one amino acid by another in proteins, for instance, by site-directed mutagenesis. In proteins, residues are densely packed and in tight van der Waals contacts. The replacement of a residue by a bulkier one will result in destabilization

of the protein structure because a larger residue will not fit in, and the structure of the protein will be distorted. The replacement of a larger residue by a smaller one is usually also unfavorable because it produces cavities that decrease the number of protein-structure-stabilizing interactions. This can be compensated by introducing solvent molecules—usually water—into the cavity, depending on the nature of the solvent and the nature of the groups that come into contact with the solvent.

Most amino acids are amphipathic structures, that is, a combination of polar or hydrophylic, and nonpolar or hydrophobic chemical groups. The differences in the hydrophobicity of amino acid side chains have been assessed by different methods, including a measurement of the partition coefficient of the model for each side chain distributed between the vapor and the water; a comparison of the solubility of amino acids and side-chain model compounds in water, ethanol, or dioxan; a measurement of the partition coefficient between water and octanol; and others. As observed, the obtained values for the hydrophobicity measured by various methods differ considerably. In general, amino acids Val, Ile, and Leu, containing aliphatic side chains, are the most hydrophobic, and those that form strong hydrogen bonds with water are the most hydrophilic, including Arg, Asp, Glu, and Lys, which have charged side chains at the physiological pH. When amino acids are incorporated into the peptide or protein structure, a part of the hydrophilicity is lost, but the –CO- and –NH- groups that participate in the peptide bond are polar, and thus retain some of their hydrophilic character. The hydrophobic/hydrophilic properties of amino acids as parts of the protein structure are probably best represented by the hydropathy index (Table 1.1). A hydropathy scale has been introduced by Kyte and Doolittle on the basis of a large amount of data from the literature, and is based on a computer program that determines the relative occupancy in the hydrophobic or hydrophilic environment within the proteins for each amino acid. A good correlation was demonstrated by comparing the obtained values and the position of a particular amino acid in known protein structures determined by crystallography.

Hydrophobicity is a very important factor in protein stability: the hydrophobic collapse of the nonpolar residues forming the nonpolar protein core is believed to play a fundamental role in the spontaneous folding of proteins. Hydrophobic residues with a large positive hydropathy index are found in the protein interior, while polar residues with a large negative hydropathy index occur on the protein surface exposed to the water. In membrane proteins, the surfaces in contact with the inner membrane lipid layer are also composed of nonpolar residues.

1.1.2 Standard Amino Acids by the Nature of the Radical

The hydrophobic amino acids with an aliphatic side chain are Gly, Ala, Val, Leu, Ile, Met, and Pro (Figure 1.1). Gly is the smallest amino acid, with just one hydrogen atom as a radical, which does not provide the molecule with a pronounced hydrophobic nature. Gly is also not chiral, and is therefore sometimes classified as a special amino acid not belonging to any particular group. Ala, Val, Leu, and Ile all have a short hydrocarbon side chain, which is branched, except in the case of Ala. The side chain of Ile has an extra asymmetric carbon atom. Met is a sulfur-containing amino acid that has a hydrocarbon chain and a methyl group attached to the sulfur. This makes Met a

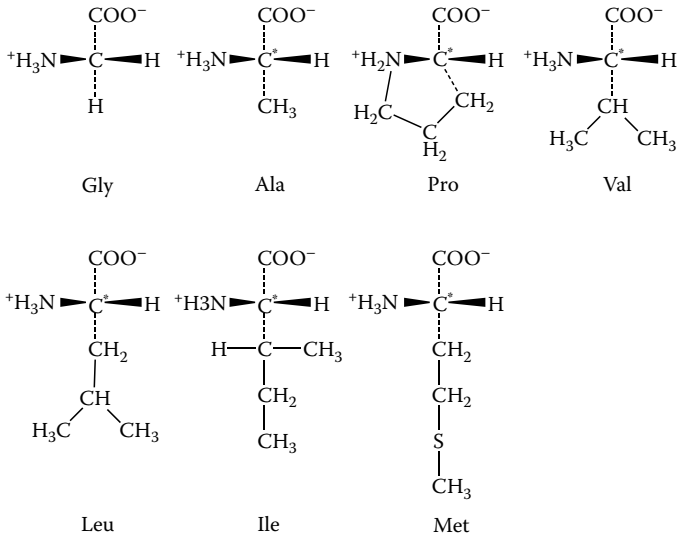


FIGURE 1.1 Structure of the hydrophobic amino acids with an aliphatic side chain.

suitable methyl-group donor in a number of methylation reactions in the cell. The sulfur atom can be oxidized by air into sulfoxide and sulfone. Although the Met side chain has a dipole moment, it also has a nonpolar and hydrophobic character. In proteins of prokaryotes, Met is the first amino acid in the sequence and also during the assembly of proteins in eukaryotes, but is later stripped away. Pro has an aliphatic but cyclic side chain; this results in special conformational properties relating to its location in proteins. The Pro ring is not totally planar; Cγ is out of the plane by about 0.05 nm.

The hydrophilic uncharged amino acids are Ser, Thr, Asn, Gln, and Cys (Figure 1.2). The polarity of the hydroxyl groups in Ser and Thr, and of the amide and carbonyl groups in Asn and Gln, ensures strong and multiple hydrogen bonding with water. Cys contains a thiol group that is related to the hydroxyl group but is chemically different—the hydrogen bonding is very weak, and the group is essentially hydrophobic, as is illustrated by the positive value of the hydropathy index (Table 1.1). However, it is able to ionize, and this is the reason why Cys is classified in this group. Ser is an important precursor in the biosynthesis of purines; pyrimidines; several amino acids, including Gly, Cys, and Trp; and other metabolites, including sphingolipids and folate. Ser is included in the active site of many enzymes, particularly in hydrolases, where it plays an important catalytic role as a nucleophile reagent. It is also the site of the glycosylation and phosphorylation of proteins. Ser is important in the mechanism of many kinases, and its phosphorylation modulates the activity of many proteins and is the control point of many cell-signaling cascades. Thr also contains the hydroxyl group in its side chain, and this results in an additional side chain chiral center. Its role in posttranslational modifications, kinases, and signaling is similar to that of Ser. Asn and Gln are the amidated forms of Asp and Glu, respectively, with side chains that do not ionize and are not very reactive. However, they do form strong hydrogen bonds and are used in proteins for

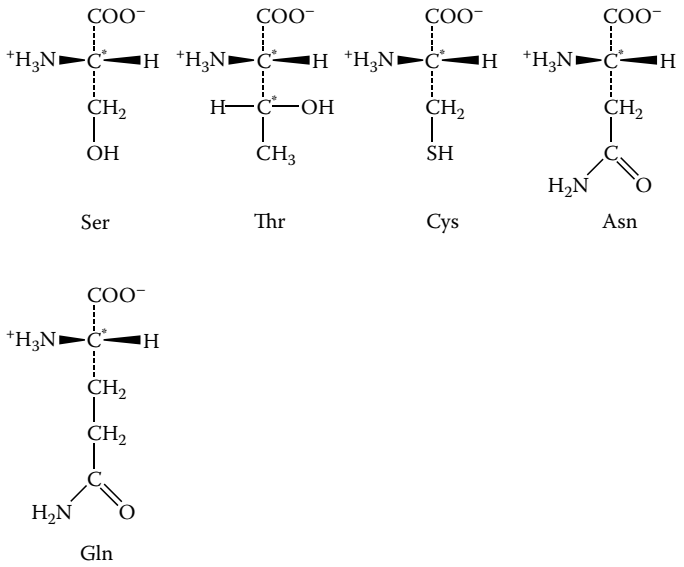
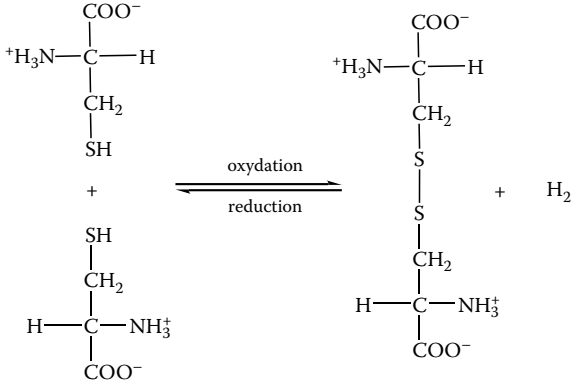


FIGURE 1.2 Structure of the hydrophilic amino acids with an uncharged side chain.

stabilizing the structure, for protein–protein interactions and for protein–DNA interactions (see Chapter 22, Figure 22.4). At very high and very low pH values and at high temperatures, Asn and Gln can be converted to Asp and Glu. Cys is part of the active center of the cysteine proteases, where, in general, it has the same role as Ser in hydrolases. The thiol group of Cys is relatively reactive, and can be alkylated by alkyl halides and added to double bonds. It also interacts with metal ions and forms complexes, as in zinc fingers (see Chapter 22). It is the target group in heavy-metal poisoning, particularly with the Hg^{2+} ion. A very important property of Cys is its ability to form disulphide bridges that play a key role in stabilizing the structure of many proteins.



(1.3)

Disulphide bonds formed by Cys residues have a nonpolar surface area.

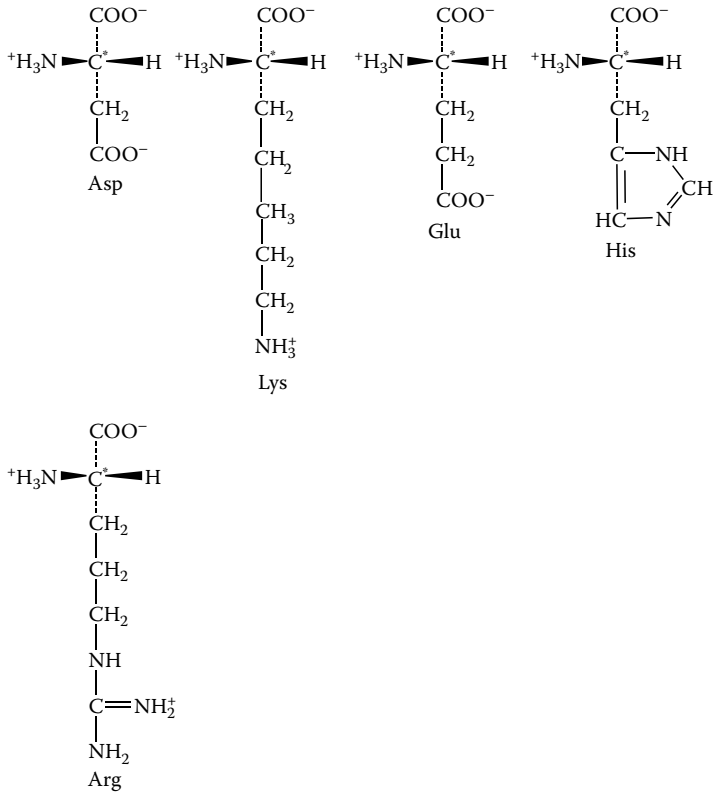
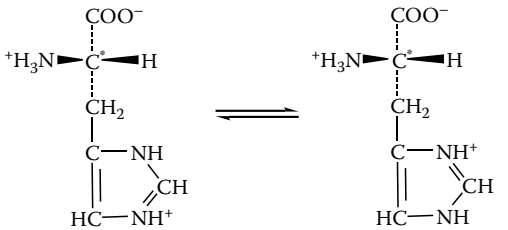


FIGURE 1.3 Structure of the hydrophobic amino acids with a charged side chain.

The charged amino acids are Lys, Arg, His, Asp, and Glu (Figure 1.3). The positively charged Lys, Arg, and His have an amino (or imino in His) group in the radical and negatively charged Asp, and Glu has a carboxyl group. Their charged radicals interact with water and other molecules by charge-to-charge interactions and hydrogen bonding. They are mostly found on the surface of the proteins. Lys is very basic and relatively reactive, particularly at higher pH values, where it is not charged. The important reaction of the Lys side-chain amino group is the formation of a covalent Schiff base with aldehydes, which is used to link the polypeptide chain with another one, for example, in collagen, and with other aldehyde groups containing molecules, for example, with pyridoxal phosphate in many enzymes. In proteins, it is the target for a number of posttranslational modifications, including methylation, acetylation, ubiquitination, hydroxylation, and glycosylation. ε-NNN-trimethylated Lys is a precursor of carnitine, which is involved in fatty-acid transport across the mitochondrial membrane. Lys is used for the treatment of the Herpes simplex infection because this virus needs Arg for its replication, and a supplement of Lys decreases Arg's availability. Arg has a planar guanido group in which the carbon atom is in the sp^2 hybridization state. The guanido group has the ability to form multiple hydrogen bonds and to

attract negatively charged groups, both inside the protein structure and in the interaction of proteins with other molecules. A notable example is the interaction of Arg in DNA-binding proteins with DNA base pairs (see Chapter 22, Figure 22.4). The guanido group is also the site of methylation, which can regulate several processes in the cell, including the binding of DNA and RNA to the DNA- and RNA-binding proteins, leading to a modified transcription and RNA processing, and signal transduction by alternating protein–protein interactions. Arg also functions as a precursor of nitric oxide (NO), urea, an important regulator of blood pressure, by which nitrogen is excreted from the bodies of humans and many other organisms, and creatine, from which creatine phosphate, the energy-supplying molecule in muscles, is produced, as well as several other molecules. Because of its involvement in the production of NO, in the release of growth hormone, and in other processes, Arg supplements are marketed, though their beneficial effect is controversial. His has an imidazole ring in which two nitrogen atoms have different characters: one is slightly acidic and the other is basic, which can bind a proton. Because of the resonance structure of the ring, the location of the basic nitrogen is not fixed:



(1.4)

In enzymes, His is utilized in catalytic triads (see Chapter 21, Equation 21.3 and Figure 21.6) and to shuttle protons, for example, in carbonic anhydrase. In some proteins, it is used to form coordinate covalent bonds with metal ions; a well-known example is the zinc fingers in DNA-binding proteins (see Chapter 22). It is also a precursor of hormone histamine. His is the only amino acid with pK_R around the physiological pH (Table 1.1) and gives the proteins the properties of a buffer if it is present in considerable quantities, for example, in hemoglobin. Asp and Glu are both acidic amino acids that are present in the catalytic triad of hydrolases (see Chapter 21, Equation 21.3 and Figure 21.6). The shorter distance between the α-carbon atom and the carboxylic group in the radical in Asp than in Glu makes Asp the stronger acid (Table 1.1) and thus more efficient in the triad. The carboxyl groups of Asp and Glu have the same chemical properties as other organic acids, including acetic acid. These groups are used by proteins and peptides to bind metal ions, most notably Ca^{2+} and Zn^{2+}. Asp and Glu participate in the biosynthesis of several amino acids, including Ala, Met, Thr, Ile, and Lys. Glu has a key role in amino-acid transamination, an important step in amino-acid degradation and synthesis. Together with the deamination of Glu, transamination is involved in nitrogen excretion via urea. Glu is also an important neurotransmitter in mammals and other species and a precursor of GABA (γ-amino butyric acid), the main inhibitory neurotransmitter in the mammalian central nervous system. Glu can also be used as a food additive to enhance taste and flavor.

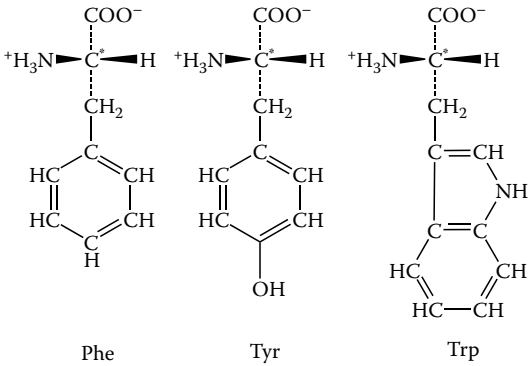


FIGURE 1.4 Structure of the amino acids with an aromatic side chain.

The aromatic amino acids are Phe, Trp, and Tyr (Figure 1.4), all of which are bulky and essentially hydrophobic, but the delocalized π-electrons in the ring systems can take part in weak electrostatic interactions (see Chapter 2). They all absorb ultraviolet light, most notably Trp. The molar extinction coefficients of Trp, Tyr, and Phe at the local maximum 280 nm and pH 7 are around 5500 cm^{-1}/M, 1200 cm^{-1}/M, and 200 cm^{-1}/M, respectively. This property is used to detect the proteins in solution. Trp, Tyr, and Phe are also the source of the intrinsic fluorescence in proteins. The quantum yield of the emitted light after excitation at 280 nm is 0.20 for Trp (emitted at 348 nm), 0.14 for Tyr (emitted at 303 nm), and 0.04 for Phe (emitted at 382 nm). The emission is very sensitive to the fluorophore surroundings, and this is employed to follow the changes in protein structure (see Chapter 11). The Tyr side chain is mainly hydrophobic, with the exception of the –OH group, which is capable of hydrogen bonding and dissociation at high pH (Table 1.1). This group is also the site of sulfonylation and phosphorylation, the latter being an important event in signal transduction via receptors with tyrosine kinase activity (see Chapter 23). Tyr is a precursor of the adrenal hormones dopamine, epinephrine, and norepinephrine, and the thyroid hormones triiodothyronine and thyroxin. Dopamine is also synthesized in various regions of the brain, where it functions as a neurotransmitter. Trp with an indole side chain is the largest amino acid. The nitrogen in the ring can participate in hydrogen bonds as a hydrogen donor; in other cases, the ring is hydrophobic. Trp is a precursor of the neurotransmitter serotonin and the neurohormone melatonin.

For the detection and identification of individual amino acids, mass spectrometry, chromatography, isoelectric focusing, optical and fluorescence spectroscopy (aromatic amino acids), and different combinations of these methods can be used. The amino group of amino acids reacts with ninhydrin, giving a colored product, and reacts with fluorescamine, giving a fluorescent product. A fluorescent derivative of the amino acids is also obtained in the reaction with dansyl chloride; however, this reagent is not specific just to the amino group but will also react with some other groups, including the hydroxyl group in phenols. These three classical reactions to detect amino acids are shown in Figure 1.5. Some of the reactions specific to the side chains of amino acids are

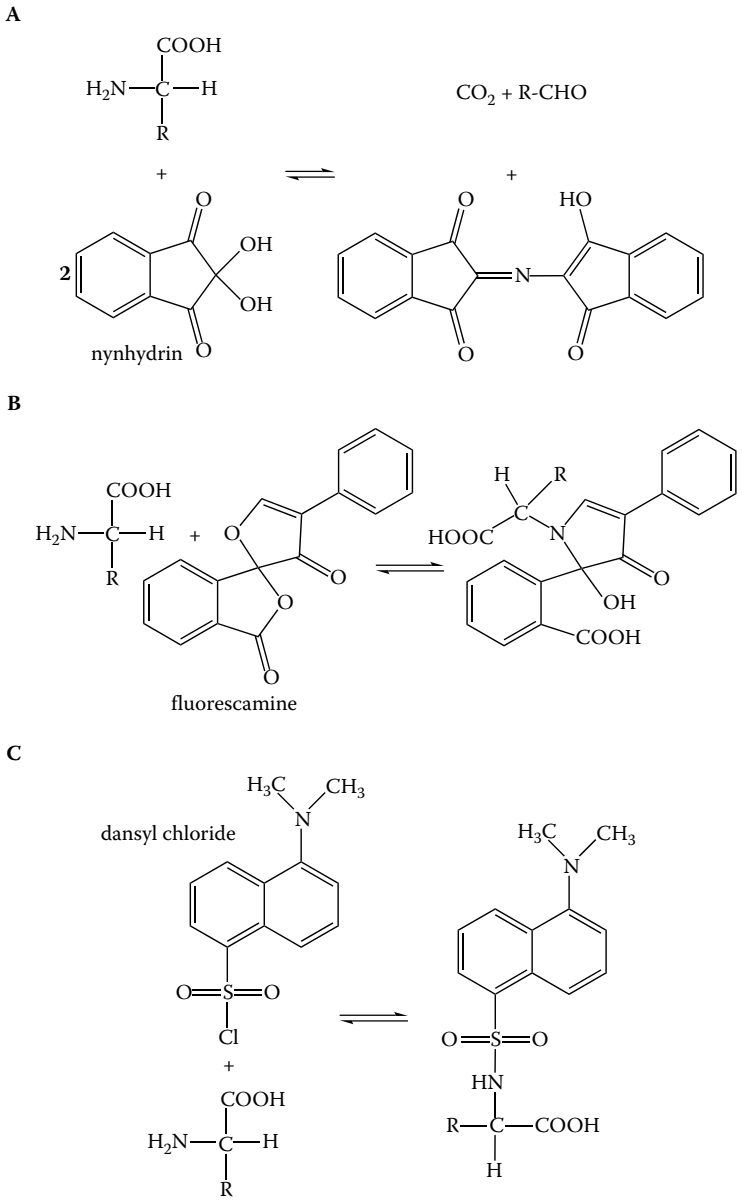


FIGURE 1.5 Reactions for the detection of amino acids by spectrophotometry and fluorescence: (A) ninhydrin reaction; (B) reaction with fluorescamine; (C) reaction with dansyl chloride.

- Reactions with Elman reagent (5.5′-dithio-bis-(2-nitrobenzoate); 2,2′-dipyridyl disulfide; and iodoacetic acid or iodoacetamide for the thiol group in Cys)
- Reactions with tetranitromethane; $HgNO_3$ (Millon reaction); phosphomolybdotungstic acid (Folin–Ciocalteu reaction); boiling concentrated nitric acid (xanthoproteic reaction); and α-nitrous-β-naphtol for the hydroxyl group in Tyr (the xanthoproteic reaction can also be used to detect Trp and Phe)
- The reaction with glyoxylic acid in H_2SO_4 (Hopkins–Cole reaction) for Trp
- The reaction with α-napthol and sodium hypochlorite (Sakaguchi reaction) for Arg

Most of these reactions give colored products that are suitable for spectrophotometric detections.

1.2 NONSTANDARD AMINO ACIDS

A vast number of other amino acids, in addition to the standard 20 acids, can be found in peptides and proteins as well as free in cellular and extracellular solutions. The number of known nonstandard amino acids exceeds 700, and this number continues to grow. The role of many of them is not well understood. However, nonstandard amino acids can be divided into two groups: amino acids that differ from standard amino acids in terms of the structure of the radical, and D-amino acids, which are structurally equal to the standard amino acids but are of different chirality. It seems that those found in peptides and proteins, synthesized via ribosomal machinery, are all converted into the nonstandard form from the standard amino acids after polypeptide chain synthesis. These posttranslational modifications will be covered in more detail in Chapter 5. Most modified amino acids are obtained by the attachment of additional functional groups to the radical. Important examples of this type of nonstandard amino acids are hydroxylated Pro and Lys; phosphorylated Ser, Thr, and Tyr; methylated Lys, His, and Arg; carboxylated Glu, and many others (for some structures, see Figure 1.6). The only known exception to this rule is the occurrence of selenocysteine in a few proteins, including the enzymes glutathione peroxidase, glycine reductase, and some hydrogenases (Figure 1.6). Selenocysteine is included in the ribosomal biosynthesis of proteins by using special selenocysteine-binding tRNA that recognizes the selenocysteine insertion sequence and the UGA codon (normally the stop codon) in mRNA. Selenocysteine is sometimes regarded as the 21st coded amino acid. The synthesis of selenocysteine is achieved by enzymatically catalyzed complex modifications of Ser that are attached to the tRNA. Mechanisms to convert Ser to selenocysteine and the required enzymes differ in terms of prokaryotes and eukaryotes.

Some peptides are synthesized in the cells nonribosomally by a sequence of enzyme-catalyzed reactions facilitating the incorporation of nonstandard amino acids and also non-amino-acid building blocks such as carboxy acids, heterocyclic rings, and fatty acids. Most of these peptides are synthesized by prokaryotes. The peptides in this category frequently include D-amino acids, and many of them are cyclic, for

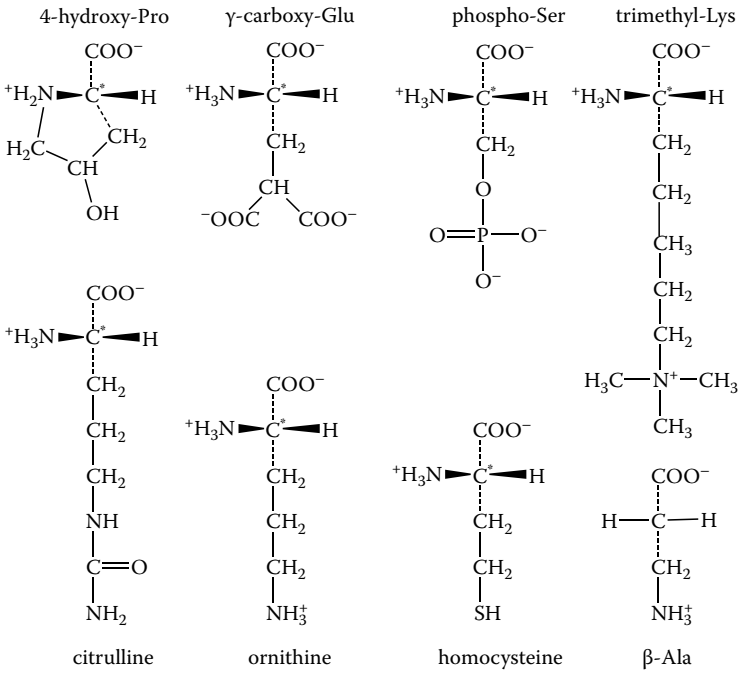


FIGURE 1.6 Some of the nonstandard amino acids.

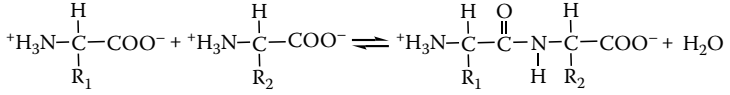
instance, the peptide antibiotics valinomycin, gramicydin, and actinomycin D, and cyclosporin A. Microorganisms also synthesize other classes of antibiotics in which the conversion of L-amino acids into the D-form is a part of the posttranslational modifications, for example, in the so-called lantibiotics epidermin and cinnamycin. Higher organisms also synthesize D-amino-acid-containing peptides, but in these peptides the standard L-amino acids are in all known cases converted to the D-form after the synthesis is complete. Examples include muscle-contracting peptides such as achatin I from snails, opioid peptides such as demorphins, and antimicrobial peptides such as bombins from the skin of certain frogs and toads, as well as the peptide toxin from the platypus. D-amino acids are components of bacterial cell walls containing D-Ala and derivatives of D-Glu, while D-Tyr is a constituent of the wall of yeast spores. D-amino acids are also found in plants, for example, in some alkaloids such as N β-(D-Leu-D-Arg-D-Arg-D-Leu-D-Phe)-naltrexamine. Another nonstandard amino acid is β-Ala (Figure 1.6), common in the naturally occurring peptides carnosine (β-Ala-His) and anserine (β-Ala-methyl-His), and also in pantothenic acid (vitamin B_5), which is itself a component of coenzyme A. There seem to be two main functions of nonstandard amino acids in peptides and proteins. The first is to protect the polypeptide chain against decomposition by the proteolytic enzymes that are not able to recognize the unusual amino acids. The second function is to provide additional functional groups for special purposes. A number of nonstandard amino acids that have never been identified as peptide and protein constituents have been found in organisms. Known examples in humans include homocysteine, ornithine

and citrulline, as well as D-Ser and D-Asp, and others. The first three (Figure 1.6) are important intermediates in metabolic processes, ornithine and citrulline are involved in the biosynthesis of urea, and homocysteine takes part in amino-acid metabolism. N-Methyl-D-Asp and D-Ser are the agonist and coagonist of the N-Methyl-D-aspartate receptors involved in excitatory glutamatergic synaptic transmission. D-Amino acids are of importance in peptidomimetic drugs, where their introduction into the structure enhances the resistance to proteolysis and increases the bioavailability of the drug (see Chapter 31). Peptide drugs based on D-amino acids that effectively inhibit the entry of the AIDS virus HIV-1 into cells are in common use today.

One obstacle to knowing more about the occurrence of D-amino acids in peptides and proteins is that they are not easily identified. Recently, Sweedler and coworkers have developed a method to analyze for them in natural peptides. The method is based on the selective degradation of the polypeptide chain with microsomal alanyl aminopeptidase, combined with mass spectrometry. The enzyme selectively degrades the peptides that lack D-amino acids. By comparing a sample before and after digestion, the D-amino-acid-containing peptides can be identified even when they are present in very small quantities. This approach should help to increase the number of detected D-amino acids in natural peptides and proteins.

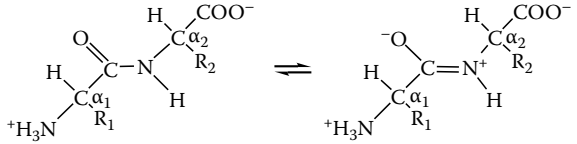
1.3 THE PEPTIDE BOND

Polypeptides are made by binding amino acids into linear polymers via peptide bonds. A peptide bond is a covalent bond that is formed in a condensation reaction between the carboxyl group of one amino acids and the amino group of the other, releasing, as a result, a molecule of water. The peptide bond is a form of the amide bond:



(1.5)

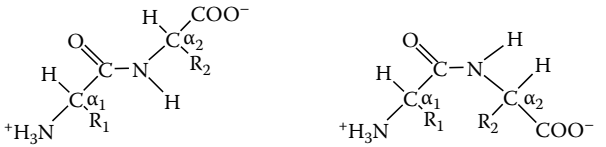
The free-energy change for this reaction is positive ($\Delta G \approx 10$ kJmol^{-1}). The equilibrium of this reaction is shifted toward the reactants, meaning that the peptide bond has a tendency to be hydrolyzed. However, the activation energy of the hydrolysis is very high (80–130 kJmol^{-1}, depending on the participating amino acids and other factors), and the decomposition of the peptide bond is thus very slow. The half-life of the peptide-bond hydrolysis at pH 7.0 and 25°C is measured in years, and the bond has properties of a partial double bond due to the delocalization of an electron pair from the nitrogen to the carbonyl oxygen:



(1.6)

The two structures shown in Equation 1.6 are in resonance, and an approximately 40% double-bond character is assumed. The partial double-bond character has

several consequences. First, the bond is planar, and rotation around the C–N axis is impossible. This forces six atoms—the C and N participating in the bond and all the four atoms connected to them (O, H, $C\alpha_1$, and $C\alpha_2$; see Equation 1.5)—to be located on the same plane. Second, the partial double bond can be in two configurations—trans and cis:



(1.7)

The preferred configuration is trans, and the ratio of cis to trans was found to be about 1:1000 in peptides and proteins, except when the Pro is on the N side of the peptide bond, in which case the ratio of cis to trans is 1:20. The trans configuration is favored because, in this configuration, the overlap and repulsion of the radicals is minimized. In the case of Pro, in which the peptide-bond-forming nitrogen atom is a part of the ring structure, the overlap of radicals in the cis configuration is considerably decreased, resulting in about 10% of these peptide bonds in proteins being cis. In Pro, the lack of a hydrogen atom attached to the nitrogen of the peptide bond that could help in the resonance stabilization of this bond also corroborates the increased occurrence of the cis configuration. However, the ratio of cis to trans is also dependent on other factors, and is considerably lower in more flexible and less structured small peptides, where up to 30% of proline-following peptide bonds can be in the cis configuration. Another consequence of the partial double-bond character of the peptide bond with a C–N spacing of 0.132 nm is that this bond is shorter than the single C–N bond (a C–N distance of 0.145 nm) and longer than the double C–N bond (a C–N distance of 0.125 nm). The distances and angles between the atoms in the peptide bond are shown in Figure 1.7. An additional important feature of the peptide

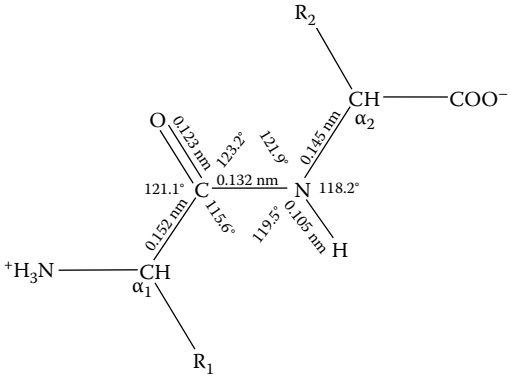
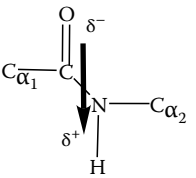


FIGURE 1.7 Distances and angles between the atoms in the peptide bond.

bond is the polarity. Because of the delocalization of a nitrogen electron pair and the differences in the electronegativity of the participating atoms, the peptide bond shows a dipole moment of around 3.5 Debye units:



(1.8)

Finally, the partial double-bond character of the peptide bond results in the absorbance of light in the UV region at wavelengths between 190 and 230 nm. The properties of the peptide bond are important for the structure of peptides and proteins, as shown, for example, in Chapter 3.

FURTHER READING

1. Barrett, G. C. 1985. Chemistry and Biochemistry of Amino Acids. New York: Chapman and Hall.
2. Amino Acids, Peptides and Proteins. Specialist Periodical Reports of the Royal Society of Chemistry, Volumes 1–30.
3. Wolosker, H., Dumin, E., Balan, L., and Foltyn, V. N. 2008. D-Amino acids in the brain: D-serine in neurotransmission and neurodegeneration (Minireview), FEBS J. 275, 3514–26.

2 Noncovalent Interactions

Matjaž Zorko

CONTENTS

2.1 The Electrostatic Interaction 21
2.2 The van der Waals Interaction 23
2.3 The Hydrogen Bond 27
2.4 The Hydrophobic Interaction 28
2.5 Multiple Interactions and Cooperativity 31
Further Reading 34

The polypeptide backbone is formed by covalent peptide bonds, as presented in Chapter 1. All the higher structures of peptides and proteins arise from the interactions of amino-acid residues protruding from this backbone. These interactions are as follows: the electrostatic interaction, the van der Waals interaction, the hydrogen bond, the hydrophobic interaction, and the –S-S- bridges between Cys residues. All these interactions—with the exception of the covalent –S-S–bridges—are weak and strongly affected by their environment. However, the interactions are numerous and do not only stabilize the physiologically functional native structure of the proteins and peptides but also provide the basis for their dynamic behavior.

2.1 THE ELECTROSTATIC INTERACTION

Electromagnetic interactions provide the fundamental basis for all the different bonded and nonbonded interactions between atoms and molecules. We will discuss here the interactions between charged atoms, functional groups, and molecules. According to Coulomb's law, the force between two interacting charges A and B can be described by the equation

$$F = \frac{q_1 \cdot q_2}{D \cdot d^2} \tag{2.1}$$

where q_1 and q_2 are charges, d is the distance between the charges, and D is the relative dielectric constant of the medium in which the charges are situated. The values of the dielectric constants of some common solvents are listed in Table 2.1. In crystals the value of D is considered to approach 1—the value in vacuum—and

TABLE 2.1
Relative Dielectric Constants of Some Common Solvents

Solvent	Dielectric Constant	Solvent	Dielectric Constant
Vacuum	1	Dichloromethane	9.1
CCl_4	2.23	Dimethylformamide	36.7
Benzene	2.27	Water	80
Ethylacetate	6.02	Formamide	109

the electrostatic interactions, also called ionic bonds, in ionic crystals are strong, the energy of the bond being over 100 kJ mol^{-1}. In a protein interior, the value of D is 2–4, and the interaction is also relatively strong; however, because of solvation, charged groups do not tend to enter the protein interior, and such interactions are rare. Although strong, the internal electrostatic interactions actually destabilize the protein structure because the solvated state of the charged groups in an unfolded protein is more thermodynamically favorable than the nonsolvated state in the interior of the folded protein. In an aqueous environment with a high dielectric constant of around 80, the electrostatic interaction becomes a weak interaction. This is due to the shielding effect of water dipoles enclosing charged ions and functional groups (Figure 2.1). Most of the electrostatic interactions between charged amino-acid residues—Asp, Glu, Lys, Arg, and His at the physiological pH (see the corresponding values of pK in Chapter 1, Table 1.1)—occur mainly on the protein or peptide surface in close contact with water, and are, as a result, weak. For instance, a positively charged Lys or Arg residue can interact with negatively charged side chains of Asp or Glu. In proteins this interaction is referred to as a salt bridge. As observed in the solved three-dimensional structures of proteins, salt bridges are relatively rare

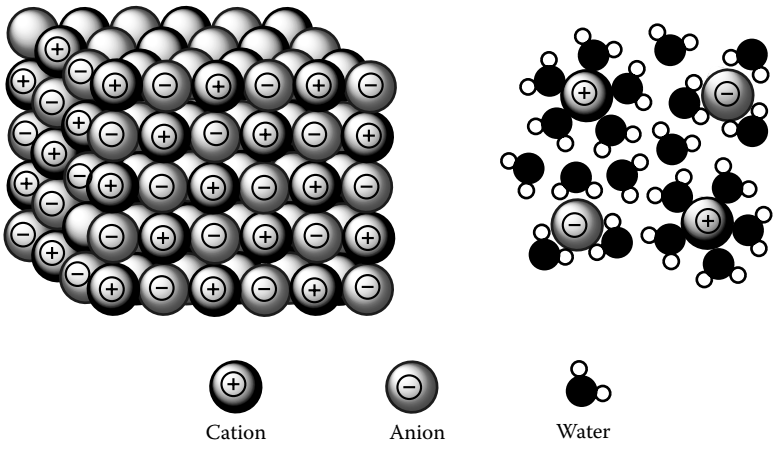


FIGURE 2.1 Electrostatic interactions in a crystal and in a water solution.

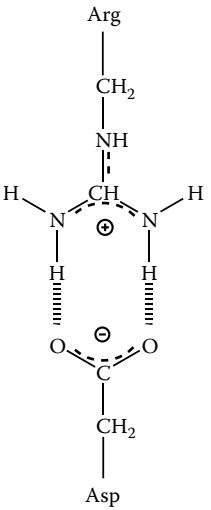


FIGURE 2.2 Interaction of Arg with a carboxylic group.

in such material, and they normally occur on the surface. An exception is when an internal salt bridge is involved in the catalytic mechanism of an enzyme such as in the Asp–His–Ser triad of serine proteases (see Chapter 21.) An additional problem of salt bridges in proteins is the delocalization of charge. Coulomb's law is invariable for point charges; however, in amino-acid residues the charges can be distributed across the broader surface of the residue, and the interaction is thus weakened. One such example is the residue of Arg, where the positive charge is shared by three amino groups. When Arg is interacting with the carboxyl group of Asp or Glu, where the negative charge is also delocalized and shared by both carbonyl oxygens in the ionized group, the electrostatic interaction is much weaker than in the corresponding hydrogen bonds that are also formed between these two groups (see Figure 2.2). Salt bridges on the protein surface encounter another problem: they are situated on the border between the less dielectric protein interior and the more dielectric solvent surroundings. Two charges on the spherical protein surface have the shortest mutual distance via the protein interior but, as was observed, they interact across the surface via a high dielectric solvent, which makes the interaction much less strong. Bulky, nonpolar groups in the vicinity of the charge and with limited solvent accessibility can further reduce the interaction energy.

2.2 THE VAN DER WAALS INTERACTION

Two opposite forces operate between two atoms approaching each other: the repulsive and the attractive van der Waals forces. The magnitude of each of these two forces depends differently on the interatomic distance. With a decreasing interatomic distance, the repulsive force increases much faster than the attractive one, as shown in Figure 2.3. The resulting curve is the sum of both forces and shows two regions of net interaction: the attractive and the repulsive. It is hard to give a precise

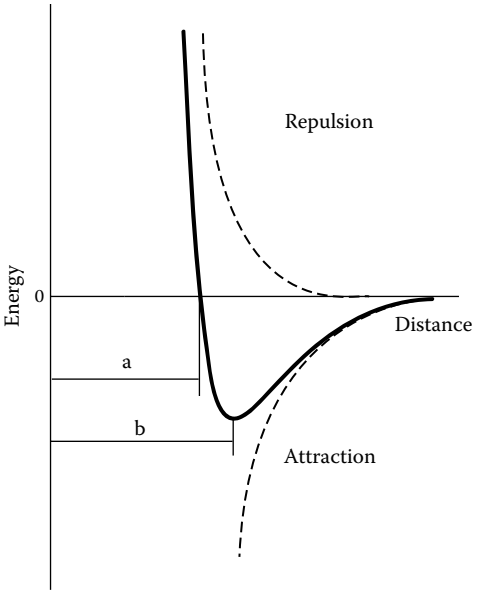


FIGURE 2.3 Schematic presentation of the van der Waals interaction—repulsive and attractive forces as the function of distance. Label a is the sum of the van der Waals radii; b is the distance between the centers of the two atoms in the optimal van der Waals interaction.

mathematical presentation of the resulting curve since any generalization will lead to an approximation. The net potential energy of the interaction is usually given as the Lennard–Jones potential:

$$E = \frac{C_n}{d^n} - \frac{C_6}{d^6} \tag{2.2}$$

where C_n and C_6 are constants, and d is the distance between the centers of the participating atoms. The first term represents repulsion, and the second term attraction. Since this is an approximation, n can be chosen arbitrarily to some extent, and a value of 12 is usually taken for n, for mathematical reasons. As a consequence, the Lennard–Jones potential is frequently referred to as the 12−6 potential.

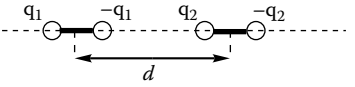
The interatomic distance where the attractive and repulsive forces are equal to one another is called the sum of the van der Waals radii. If measured for the diatomic molecule of an element, usually in crystals, the van der Waals radius for a particular atom can be obtained. For the atoms of the main bioelements H, O, C, N, P, and S, it ranges from 0.14 nm for a hydrogen atom to 0.2 nm for a sulfur atom. A value of 0.19 nm is commonly obtained for carbon. These values are approximate and dependent on the system in which they are measured. Another distance to be considered is the optimal van der Waals contact distance, defined by the minimum of the curve in Figure 2.1. The van der Waals radius defines the size of an atom as modeled by a hard-shell sphere. Squeezing two atoms below the sum of the van der Waals radii results in a large energy cost as the electron orbitals start to overlap. The Pauli

Exclusion Principle states that no two electrons can share the same state, so half the electrons of the system would have to go into orbitals with an energy higher than the valence state. The repulsive force prevents matter from collapsing and defines the size and shape of all molecules, including amino acids, peptides, and proteins. The van der Waals volumes of the coded amino acid residues are given in Chapter 1, Table 1.1.

Attractive van der Waals interactions are usually divided into three classes according to the degree of polarization of the participating molecules or their parts: (1) between permanent dipoles, (2) between a permanent dipole and an induced dipole, and (3) the dispersion interaction. Molecules or parts of molecules having a nonequal partition of charges due to the different electronegativities of the atoms are called dipoles. The separation of charges in a molecule determines its dipole moment μ, which is a vector defined by the magnitude of the separated excess charge Z and the distance d by which it is separated:

$$\vec{\mu} = Z \cdot \vec{d} \tag{2.3}$$

The dipole points from the negatively charged atom toward the positively charged one. Molecules composed of atoms with widely different electronegativities will tend to form permanently charged ions that are held together by the electrostatic interactions already discussed. When the difference in the electronegativity is moderate, the electrons in the bonds are unevenly distributed so that the centers of the negative and the positive charges do not coincide and the molecule is a permanent dipole. The permanent dipoles that are of particular relevance to polypeptide chemistry are the peptide bond, the polar groups in amino acid radicals, and water. Dipole–dipole interactions between freely rotating polar molecules tend to cancel out because the attraction between poles of different charges is compensated by the repulsion of poles with the same charge. However, the electric fields of the dipoles will not, in reality, allow a totally free rotation, and the energy of the interaction is proportional to the inverse sixth power of the distance between the dipoles, as shown for the attractive portion of Equation 2.2. When polar molecules or groups are not free to rotate, as in the amino acid residues of peptides or proteins, the energy of the dipole–dipole interaction is a complicated function that includes the dipole moments of both dipoles as well as the distance and the angle between them. In a simple case, when the dipoles are aligned as below,



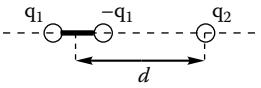
(2.4)

the potential energy of the interaction in vacuum is

$$E = -\frac{\mu_1 \cdot \mu_2}{2\pi D_0 \cdot d^3} \tag{2.5}$$

where μ_1 is the dipole moment of the dipole $q_1/-q_1$, μ_2 is the dipole moment of the dipole $q_2/-q_2$, D_0 is the dielectric constant in a vacuum, and d is the mean distance

between the dipoles. Similarly, the interaction between the dipole $q_1/-q_1$ and the point charge q_2, an approximation of the dipole–ion interaction, in the aligned positions



(2.6)

is

$$E = -\frac{\mu_1 \cdot q_2}{4\pi D_0 \cdot d^2} \tag{2.7}$$

Another type of van der Waals interaction occurs between permanent dipoles and neutral molecules in which a dipole is induced. The electric field of the ion, the charged group or the dipole, will affect the distribution of electrons in the vicinal neutral molecule, and, as a result, it will become transiently polarized. The magnitude of the induced dipole moment depends on the strength of the electric field and the polarizability of the molecule, which is a function of the structure and volume of the molecule. In general, larger molecules composed of bigger atoms with electrons that are less tightly controlled by the nuclear charges will be more affected by the external electric field than smaller ones, and will, therefore, be more strongly polarized. In principle, the induced dipole will interact with the permanent dipole or point charge in accordance with Equations 2.4–2.7, but the value of the induced dipole moment should be estimated separately, and this is usually not easy to do. It is interesting that the alignment of the induced dipole in relation to the permanent dipole is always opposite because it is defined by the orientation of the permanent dipole. The van der Waals interactions are not restricted to the polar molecules, and because of the fluctuation of the electrons in the atoms and molecules, nonpolar molecules can also transiently acquire a dipole moment, and, in turn, they can induce a dipole moment in neighboring nonpolar molecules. The interactions between induced dipoles are called dispersion interactions, or London interactions.

The strength of all forms of van der Waals interactions is in accordance with the term C_n/d^6 in Equation 2.2, but C_n is composed of different parameters and has a different magnitude in each particular type of interaction. At 25°C, the energies of the interaction of dipoles in close proximity—which are practically in contact—are comparable or smaller than the thermal kinetic energy, which is equal to 3/2RT = 3.7 kJ mol^{-1}. Although very weak and of short term, they operate within all molecules, groups, and atoms in contact and represent an important contribution to the stability of proteins and peptides.

Aromatic rings such as benzene molecules are electroneutral and are not dipoles. However, the charges in the ring are not distributed evenly. The negative charges of π-electrons are concentrated on both sides of the ring, protruding perpendicularly out of the ring plane, while positive charges are situated near the ring edge (see Figure 2.4). Molecules with this type of charge distribution are known as quadrupoles. The interaction of two aromatic structures where they lie on each other, that is, "face to face," is usual for heterocyclic structures such as Trp, particularly when several rings are stacked one above the other in a hydrophobic environment and

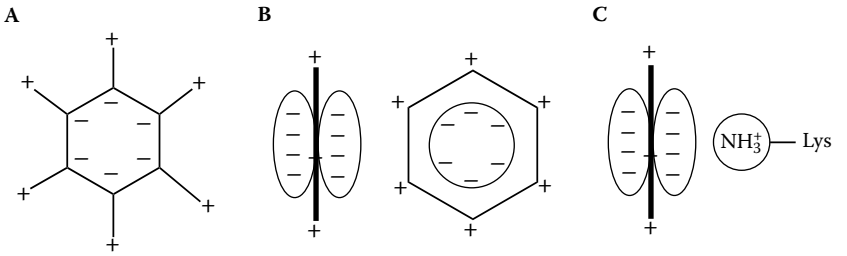


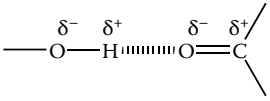
FIGURE 2.4 Interactions between aromatic rings, and the ion π interaction between an aromatic ring and an ion.

additionally stabilized by hydrophobic interactions and the resonance of adjacent π-electron rings. Studies of the crystal structure of benzene have demonstrated a "face-to-edge" packing where the edge of the ring, composed of partially positive H-atoms, interacts with the partially negative π-electrons in the aromatic structure. Phe and Tyr residues often participate in this type of interaction. The negative charge of π-electrons also attracts positively charged ions and functional groups. The residues of Phe, Tyr, and Trp, interacting with the positively charged amino group in the residues of Lys and Arg, have been identified in peptides and proteins. As determined by an analysis of these interactions in DNA-binding proteins, Arg–Phe and Arg–Tyr are the two most frequent pairs involved in cation–π interactions.

Interestingly, in DNA-binding proteins the average energy of the interaction of these pairs is around 20 kJ mol^{-1}, while in other proteins, it is approximately 40% lower. Cation–π interactions are of importance for the stability of peptides and proteins but can also be the basis of the interaction of proteins with other molecules. The neurotransmitter acetylcholine binds to the acetylcholine receptor predominantly via the interaction of its positively charged trimethylamino group with the Trp ring in the binding site of the receptor. The interaction between aromatic quadrupoles and the interaction of the quadrupole with the charged group are shown in Figure 2.4.

2.3 THE HYDROGEN BOND

The hydrogen bond can be regarded as a specific dipole–dipole interaction in which a hydrogen atom covalently attached to an electronegative atom, for example, oxygen, is approached by another electronegative atom with a lone pair of electrons, for example, carbonyl oxygen, as in the example below:



(2.8)

Being more electronegative than hydrogen, oxygen draws the electrons in the O–H bond toward itself. The hydrogen atom is then left with a net positive charge, and the oxygen is, as a result, negative. The bond between the hydrogen and the carbonyl

oxygen (Equation 2.8) is of a double nature because, in addition to the attractive electrostatic interaction between partial charges, a quasimolecular orbital between the hydrogen and the carbonyl oxygen is also formed. In the presented case, the oxygen in the –OH group is a hydrogen donor, and the carbonyl oxygen is a hydrogen acceptor. In proteins and peptides, the groups that can participate in hydrogen-bond formation are the carbonyl group in the peptide bond as a hydrogen acceptor and the –NH group in the peptide bond as a hydrogen donor, as well as the following groups in amino-acid residues: the –OH of Ser, Thr, and Tyr; the amido and carbonyl groups of Gln and Asn; the imino group in the rings of His and Trp; the carboxyl group of Asp and Glu, the amino group of Lys; and the guanido group of Arg. Groups containing hydrogen can play both the role of acceptor and of donor, whereas others can function only as hydrogen acceptors. The energy and length of the bond depend on a combination of the hydrogen acceptor and the hydrogen donor and also on the configuration of the bond. When a donor, the hydrogen and acceptor atoms are lying on the same line, and the hydrogen bond is the strongest, and it becomes weaker—though not very much—when this configuration is distorted because of the increased repulsion between the acceptor and donor atoms; the distortions can form an angle of up to 40°. The energy of the bond is between 10 and 40 kJ mol^{-1}, and the length between the donor and acceptor atoms is typically around 0.28 ± 0.03 nm. An important hydrogen-donor and hydrogen-acceptor molecule is water, which can form numerous hydrogen bonds either with other water molecules or with functional groups of the polar amino-acid residues. The ability of a molecule to form hydrogen bonds with water defines its solubility in water, and the inability to form hydrogen bonds with water is the basis of hydrophobic interactions. Hydrogen bonds are of great importance for the stabilization of the peptide and protein structures; however, they are weak enough to be broken during the conformational changes of protein molecules, thus enabling the necessary protein and peptide dynamics. Hydrogen bonding of the backbone carbonyl oxygen to the backbone amino groups leads to the formation of different secondary structures such as alpha helices and beta sheets. In addition, hydrogen bonds of the groups in amino-acid radicals contribute substantially to the stabilization of all the higher levels of the protein and peptide structures, from secondary to quaternary (see Chapter 3). Interactions can also take place between groups carrying a formal charge and hydrogen-bonding atoms; this is an especially strong variant of the hydrogen bond.

2.4 THE HYDROPHOBIC INTERACTION

In water solutions, nonpolar molecules tend to aggregate in order to minimize the amount of surface exposed to water. This process is known to be entropically, rather than enthalpically, driven. The energy needed for the ordering of nonpolar molecules in the hydrophobic interaction that keeps nonpolar molecules together is outweighed mainly by the decreased ordering of the surrounding water molecules (see Figure 2.5A). The results of a more detailed analysis of the transfer of a nonpolar cyclohexane molecule from gas, liquid, and solid phases to an aqueous environment in terms of free-energy change, and its enthalpic and entropic

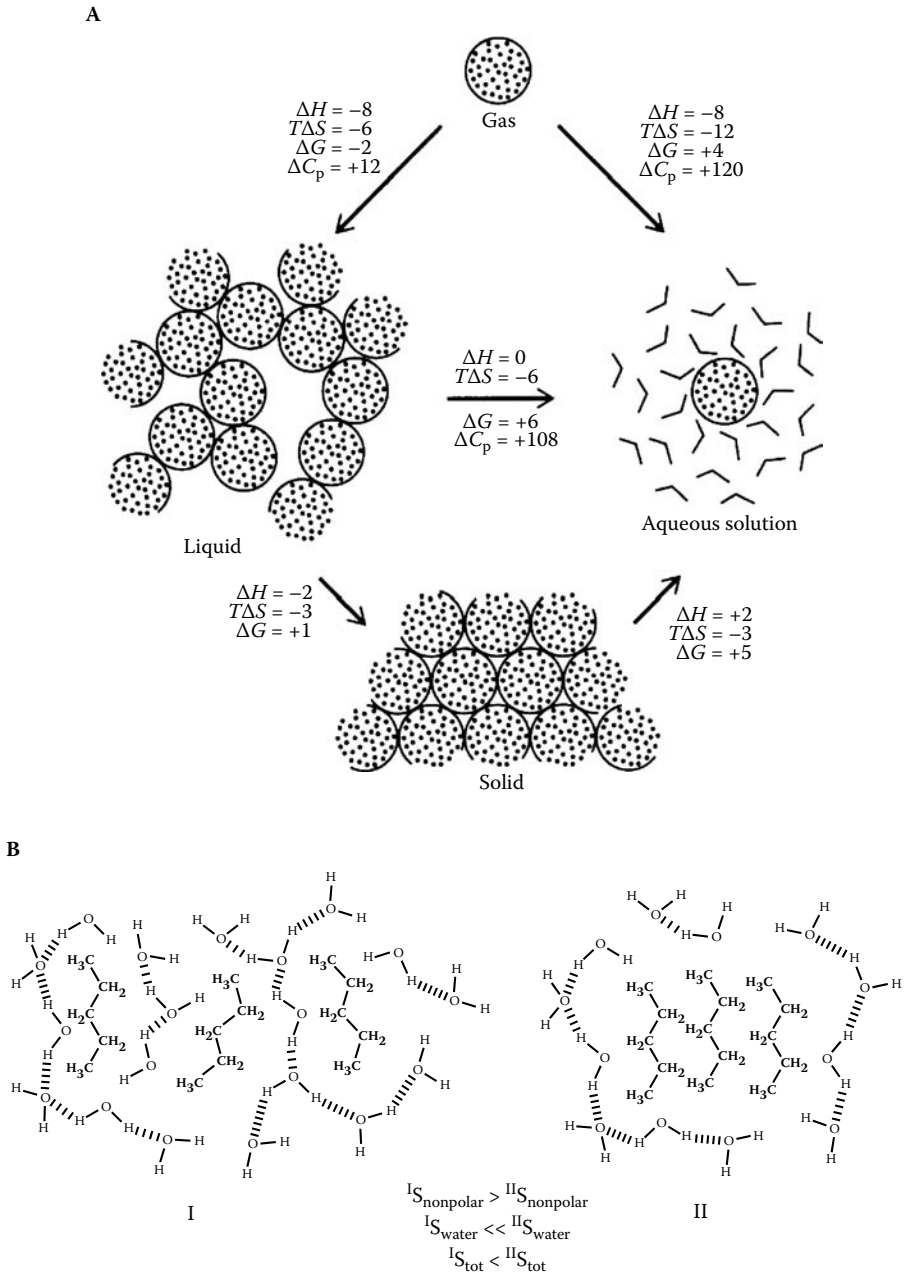


FIGURE 2.5 ΔG, ΔH, and $T\Delta S$ changes in hydrophobic interactions: (A) thermodynamic parameters for the transfer of a nonpolar molecule (cyclohexane) in water at 20°C. Units for ΔG, ΔH, and $T\Delta S$ are kcal mol^{-1}; units for ΔC_p are kcal K^{-1} mol^{-1}. (B) hydrophobic interaction is entropy driven. (Part A Reproduced from Creighton, T. E. Proteins Structure and Molecular Properties, 2nd ed., W. H. Freeman, New York, 1993. With permission.)