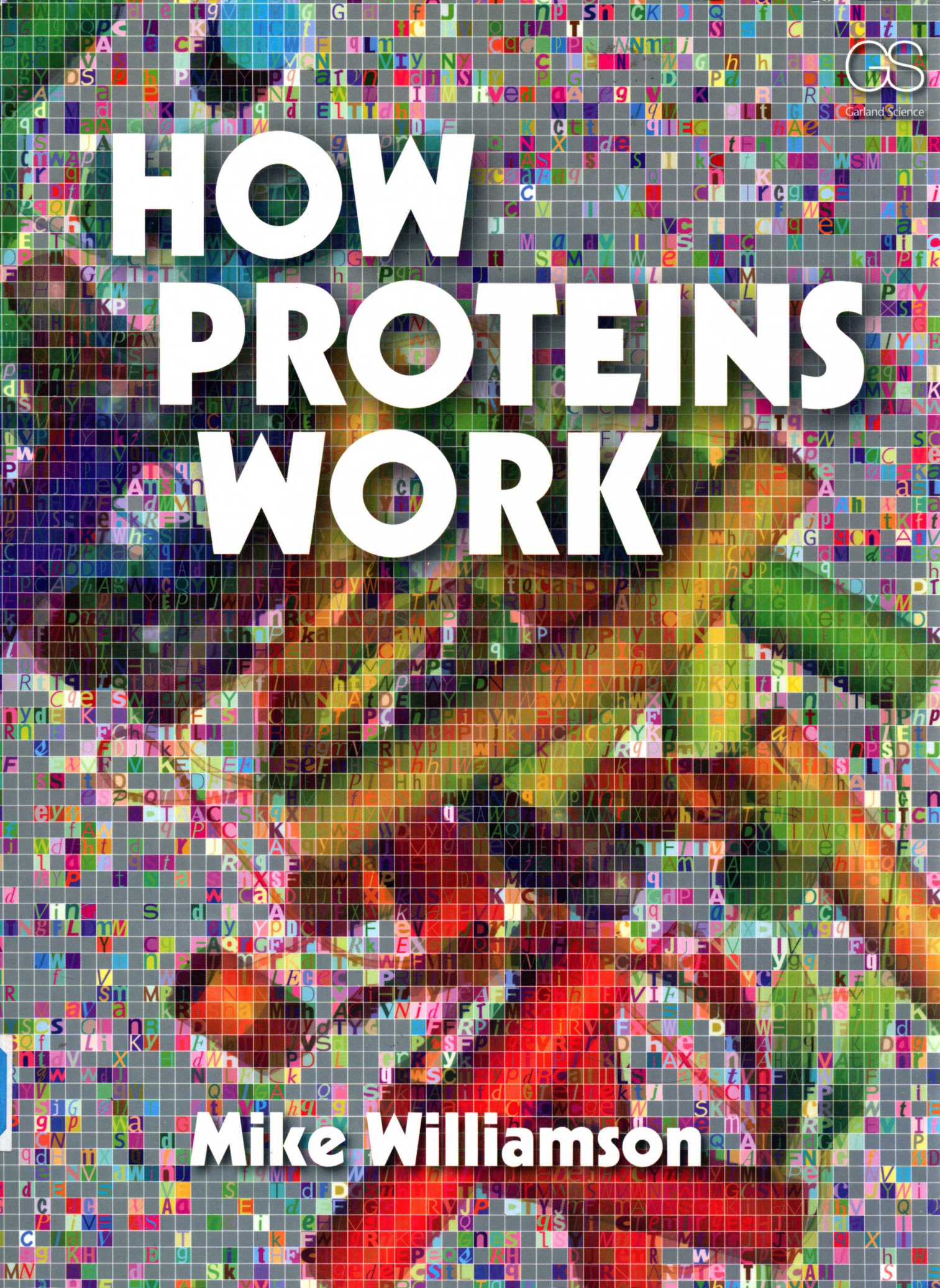

Garland Science
HOW PROTEINS WORK
Mike Williamson

HOW PROTEINS WORK

Mike Williamson

GS Garland Science
Taylor & Francis Group
NEW YORK AND LONDON

Garland Science
Vice President: Denise Schanck
Editor: Summers Scholl
Editorial Assistant: Kelly O'Connor
Production Editor and Layout: EJ Publishing Services
Illustrator and Cover Design: Matthew McClements, Blink Studio, Ltd.
Copyeditor: Bruce Goatly
Proofreader: Sally Livitt
Indexer: Medical Indexing Limited

ISBN 978-0-8153-4446-9

Library of Congress Cataloging-in-Publication Data

Williamson, Mike (Michael Paul)
How proteins work / Mike Williamson.
p. cm.
Includes bibliographical references.
ISBN 978-0-8153-4446-9
1. Proteins. 2. Proteins--Analysis. 3. Biochemistry. 4. Biophysics. I. Title.
QD431.W536 2012
572'.636--dc23
201101758

Published by Garland Science, Taylor & Francis Group, LLC,
an informa business, 711 Third Avenue, New York, NY 10017,
USA, and 2 Park Square, Milton Park, Abingdon, OX14 4RN, UK.

Printed in the United States of America

15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

GS Garland Science
Taylor & Francis Group

Visit our website at http://www.garlandscience.com

HOW PROTEINS WORK

Preface

Proteins are endlessly fascinating. They carry out almost all the catalytic functions in the cell, as well as directing and forming most of the structural framework. They catalyze reactions many orders of magnitude faster than any system that humans can devise under comparable conditions. Proteins are also much larger than most human-designed catalysts. They make many interactions of widely varying strength and duration. Crucial to any understanding of protein function is their structure: although many of the principles governing how proteins work were understood many years ago, it is not until we have the structural details that we can really appreciate exactly what the proteins are doing. This is a major reason why we don't understand membrane proteins as well as we do globular proteins. However, the structure is merely the detail that enables us to reach toward the concepts that really explain how proteins work. I have therefore endeavored to look beyond the structural detail to understand the underlying principles.

I have been occupied for some time with the study of the most essential substances of the animal kingdom: fibrin, albumin and gelatin. I conclude that the organic substance which is present in all constituents of the animal body, also as we shall soon see in the plant kingdom, could be named protein *from* πρωτειος *[proteios],* primarius, *which has the composition* $C_{400}H_{620}N_{100}O_{120}$...

Gerhardus Johannes Mulder (1802–1880)

This textbook grew out of my courses for intermediate and advanced undergraduates, and is inspired by the idea that proteins are a functional part of living and evolving systems. They have a certain form and function because it works, not necessarily as the most perfect solution to the biological problem, but certainly as a viable and successful solution. Advanced undergraduate and graduate students, as well as practitioners, interested in proteins should find the book useful. A basic foundation in chemistry and biology, as supplied by introductory undergraduate courses in biochemistry, should suffice. Students may have a background in chemistry, biology or physics, but I have tried to write the text so that it is accessible to all.

The book is written in a style that I like to read. This means the text is discursive; occasionally it goes off at a tangent; it has analogies and examples liberally scattered around; it simplifies systems as far as possible in an attempt to see the forest for the trees; and it places more emphasis on principles than on the experiments that were used to derive the principles.

There is considerable discussion of the role of evolution in tinkering with proteins to create something with a desired function. I am particularly interested in how proteins solve 'difficult' biological problems, such as catalysis, movement, and signaling. In the same way that you cannot really understand a foreign country without having some idea of its history, I believe that you cannot understand proteins unless you have some idea of how they got to their present form. The use of everyday analogies and emphasis of the physical environment around the protein enable the reader to understand proteins as well as merely know the facts about them.

Quantitative calculations are used to understand how proteins work. I strongly believe that the field of biochemistry in general, and protein science in particular, will need to place more emphasis on quantitative measurements as they mature. A holistic view—integrating structural, chemical, and biological data to try to understand how proteins help the cell to function properly—is key to this text. We are moving into a new era of biological science, where we have a good idea of many of the pieces, and we are starting to see how the pieces work together to achieve a functional whole (the idea behind Systems Biology). This book is an attempt to do exactly that.

I do not attempt to be comprehensive in the coverage of proteins. There is little coverage of medical aspects of proteins, though they are certainly described where relevant, as in signaling. I have often skipped over the experimental evidence for many of the facts presented, because I do not want to obscure the principles of how proteins work by inclusion of too much experimental detail.

Chapters 1-4 and 6 present the physical constraints that have resulted in proteins looking and working the way they do. These limitations include the structures and properties of amino acids and the forces that hold proteins together, which are discussed in Chapter 1, along with a detailed discussion of the way evolution shapes proteins. Chapter 2 discusses the domain, the fundamental structural and evolutionary building block of proteins, while Chapter 3 considers how domains associate together into oligomeric proteins; it also discusses consequences of oligomerization such as allostery and cooperativity. Chapter 4 covers an important topic that is not often discussed in textbooks, namely the cellular environment and how this influences proteins. It describes the crowded environment of the cell, how proteins bind rapidly and yet specifically to their targets, and natively unstructured proteins, as well as post-translational modifications and protein folding. Finally, Chapter 6 discusses the developing area of internal mobility within proteins.

The second half of the book, Chapters 5 and 7–10, covers various biological functions of proteins, and considers how they carry out these functions, and how their structure enables them to do so. These are enzyme catalysis in Chapter 5, movement and translocation in Chapter 7, signaling in Chapter 8, regulation (by the formation of complexes) in Chapter 9, and coordination of sequential reactions by multi-enzyme complexes in Chapter 10. Additionally, Chapter 9 looks at the results emerging from high-throughput technology. Finally, Chapter 11 discusses the techniques used in studying proteins, both experimental and theoretical.

The main text is augmented with boxes referred to by numbered asterisks (*) that provide more details on select topics, brief biographies of prominent scientists, and pedagogical analogies for further elucidation of concepts. There is also a glossary containing definitions to words that appear in bold throughout the main text.

Online Resources

Accessible from www.garlandscience.com, the Student and Instructor Resources websites provide learning and teaching tools created for *How Proteins Work*.The Student Resources Site is open to everyone, and users have the option to register in order to use book-marking and note-taking tools. The Instructor's Resources Site requires registration and access is available to instructors who have assigned the book to their course. To access the Instructor Resource Site please contact your local sales representative or email science@garland.com. Below is an overview of the resources available for this book. On the website, the resources may be browsed by individual chapters and there is a search engine. You can also access the resources available for other Garland Science titles.

For Students

Animations and Videos

The animations and videos dynamically illustrate important concepts from the book, and make many of the more difficult topics accessible.

Flashcards

Each chapter contains a set of flashcards, built into the website, that allow students to review key terms from the text.

Glossary

The complete glossary from the book is available on the website and can be searched and browsed as a whole or sorted by chapter.

Hints

The hints provide strategies and clues for solving some of the more difficult end-of-chapter problems.

Solutions to Problems

Solutions to the odd-numbered problems are provided for self-testing.

For Instructors

Figures

The images from the book are available in two convenient formats: PowerPoint® and JPEG. They have been optimized for display on a computer. Figures are searchable by figure number, figure name, or by keywords used in the figure legend from the book.

Animations and Videos

The animations and videos that are available to students are also available on the Instructor's website in two formats. The WMV formatted movies are created for instructors who wish to use the movies in PowerPoint presentations on Windows® computers; the QuickTime formatted movies are for use in PowerPoint for Apple computers or Keynote® presentations. The movies can easily be downloaded to your computer using the "download" button on the movie preview page.

Power Point Presentations

The PowerPoint presentations contain the figures and micrographs from the book. There is one presentation for each chapter.

Solutions Manual

A complete solutions manual is provided for all problems in the text.

Acknowledgments

I need to thank the many people who have helped in one way or another. These include my supervisors and mentors Dudley Williams and Kurt Wüthrich, as well as colleagues who have provided much needed insight. For advice and corrections: Pete Artymiuk and Per Bullough. For suggestions particularly on the problems: Abaigael Keegan, Hugh Dannatt, Rebecca Hill, Vicki Kent, Tacita Nye and Muhammed Qureshi. And of course the production team at Garland Science, particularly Summers Scholl who has nurtured the book through its gestation; Emma Jeffcock, who has ably managed the production process; Matt McClements, who has turned my sketches into awesome images; and Bruce Goatly, whose spot-on comments and corrections were appreciated.

Mike Williamson

Contents

Detailed Contents

CHAPTER 1

Protein Structure and Evolution

Structural biology has had an enormous influence on biochemistry in general, and on the study of proteins in particular. It can almost be said that unless we know a protein's three-dimensional structure we cannot understand how it functions. However, when the crystallographer John Kendrew determined the structure of the first protein to be described in detail (myoglobin, in 1958), the most striking feature was its irregularity and complexity (or, as Max Perutz wrote, a "hideous and visceral-looking object"—**Figure 1.1** [2]). It soon became clear that proteins require this level of complexity to bind ligands and catalyze reactions specifically. But as soon as we start looking in detail at proteins, we see that there are regular patterns to the way in which proteins fold up, patterns that are determined by the underlying structures of amino acids and by the forces that dictate how they pack together. When we look at the human body, we can identify a hierarchy of structural and functional units, each dependent on the next: limbs, organs, cells, and cellular components. The same is true of proteins—each level of structure (quaternary, tertiary, secondary, and primary) depends on the one below.

Even more importantly, the structure and function of proteins are a product of evolution. This is again true of the human body: we cannot hope to understand its functions, malfunctions, and development without understanding something about the evolutionary processes that shaped it. This is why an evolutionary viewpoint pervades this book, and why a considerable part of the first chapter has been set aside to consider the implications of evolution.

Chapter 1 lays down a framework and sets the scene for the rest of the book. It is, however, far from being just an introduction, and contains some advanced material.

The basic laws of physics can usually be expressed in exact mathematical form, and they are probably the same throughout the universe. The "laws" of biology, by contrast, are often only broad generalizations, since they describe rather elaborate chemical mechanisms that natural selection has evolved over billions of years.

Francis Crick (1988), [1]

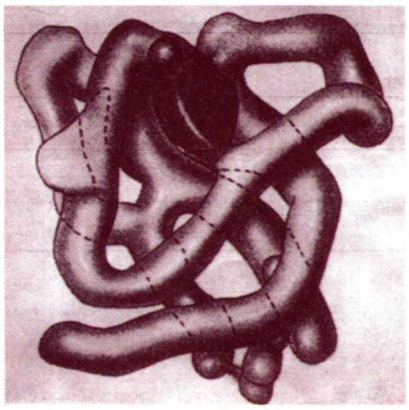


FIGURE 1.1
The first view of a protein structure was Kendrew's "hideous and visceral-looking object": the low-resolution crystal structure of myoglobin, obtained in 1958. At this resolution it is only possible to see the course of the peptide chain, much of which is in the form of α helices. Although the internal structure of an α helix is regular, the rest of the protein (the tertiary structure) is strikingly irregular. The darker region near the top is the heme, which should of course be almost completely flat; in higher-resolution structures it is indeed flat. (From J.C. Kendrew et al., *Nature* 181:662–666, 1958. With permission from Macmillan Publishers Ltd.)

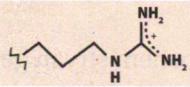
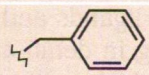
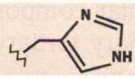
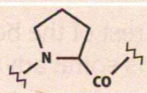
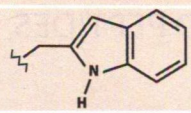
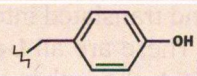
1.1 STRUCTURES OF AMINO ACIDS AND PEPTIDES

1.1.1 Proteins are composed of amino acids

There are 20 common amino acids coded for by DNA and translated into proteins from mRNA on ribosomes, as listed in **Table 1.1**. These are all **L-amino acids (*1.1)**. In addition, selenocysteine is coded for by UGA, the umber codon, which is normally a termination codon; an extra nucleotide sequence slightly downstream in the mRNA directs the cell to insert selenocysteine here. Bacteria can also produce D-amino acids and unusual amino acids by using nonribosomal synthesis, which does not concern us here but is discussed further in Chapter 10. The amino acids are known both by their three-letter abbreviations and also by one-letter codes, which match the three-letter name where possible (see Table 1.1).

An amino acid consists of a carboxylic acid, which is attached to a carbon atom called the α-carbon because it is adjacent to the carboxylate. In turn, the α-carbon is attached to an amine (hence the name *amino acid*). In the smallest amino acid, glycine, this is all there is. In all the others, the α-carbon is attached to a β-carbon, which in turn is often attached to further atoms. These are given succeeding letters from the Greek alphabet: γ, δ, etc. The carbonyl, Cα and amine are called the **backbone (*1.3)**, the other atoms being the **side chain (*1.4)**.

TABLE 1.1 The 20 common amino acids plus selenocysteine

Name	Three-letter code	One-letter code[b]	Side-chain structure[c]	pK_a of side chain	Range of pK_a in proteins	Comments
Alanine	Ala	A	CH_3			Hydrophobic, small
Arginine	Arg	R		12.5		Hydrophobic in middle, basic at end
Asparagine[a]	Asn	N	CH_2-$CONH_2$			Polar
Aspartic acid[a]	Asp	D	CH_2-CO_2^-	3.9	2.0–6.7	Acidic
Cystine/cysteine	Cys	C	CH_2-S-; CH_2-SH	8.3	2.9–10.5	Hydrophobic Reduced (SH) is called cysteine; oxidized (S–S) is called cystine.
Phenylalanine	Phe	F				Hydrophobic, aromatic
Glutamine[a]	Gln	Q	CH_2-CH_2-$CONH_2$			Polar
Glutamic acid[a]	Glu	E	CH_2-CH_2-CO_2^-	3.2	2.0–6.7	Acidic
Glycine	Gly	G	H			Hydrophobic
Histidine	His	H		6.0	2.3–9.2	Basic, aromatic
Isoleucine	Ile	I				Hydrophobic
Leucine	Leu	L				Hydrophobic
Lysine	Lys	K	CH_2-CH_2-CH_2-CH_2-NH_3^+	10.5	6.0	Basic
Methionine	Met	M	CH_2-CH_2-S-CH_3			Hydrophobic
Proline	Pro	P				Hydrophobic and hydrophilic[d]
Serine	Ser	S	CH_2-OH	14.0		Polar
Threonine	Thr	T	CH(OH)-CH_3	15.0		Polar
Tryptophan	Trp	W				Hydrophobic, aromatic
Tyrosine	Tyr	Y		9.7	6.1	Aromatic
Valine	Val	V				Hydrophobic
Selenocysteine	-	-	CH_2-SeH			Hydrophobic[e]

Amino acids have the common structure ^+H_3N–CH(R)–CO_2^-, where R is the **side chain** and the rest is the **backbone**. The table gives the structure of R.

[a]In addition, Asp and Asn are collectively called Asx with one-letter code B, and Glu and Gln are called Glx with code Z.

[b]The one-letter code for any of the 20 amino acids is usually X. The one-letter code matches the first letter of the amino acid where this is unique (C, H, I, M, S, V). Where more than one amino acid starts with the same letter, the code is assigned to the more common amino acid (A, G, L, P, T). The rest are phonetic where possible (Fenylalanine, asparagi**N**e, a**R**ginine, **Q**tamine, t**Y**rosine). Tryptophan has a double ring (double-u or W), and the others have a letter somewhere near the letter that the amino acid starts with (Asp D, Glu E, Lys K).

[c]The backbone CH carbon is the alpha carbon Cα, and its attached proton is Hα. The side-chain atoms are given succeeding letters from the Greek alphabet: β (beta), γ (gamma), δ (delta), ϵ (epsilon), ζ (zeta), η (eta). In computer files such as coordinate files, these labels are given in capital letters: A, B, G, D, E, Z, H. Where there is more than one heavy atom the same distance out from Cα, they are numbered 1 and 2; so for example the two methyl groups of a leucine are called Cδ1 and Cδ2. The dihedral angles along the side chain are called χ_1 (chi-1, pronounced kai, the angle formed by the four atoms N, Cα, Cβ, and Cγ), χ_2, and so on.

[d]The entire amino acid is drawn here. Strictly, proline is not an amino acid but an imino acid because it has an NH group, not an NH_2 group. As discussed in Chapter 4, the ring is hydrophobic but the main chain is unusually hydrophilic, making polyproline, for example, soluble in water.

[e]Selenocysteine is not normally counted as one of the standard amino acids (see the text).

*1.1 L-Amino acid

An L-amino acid is an α-amino acid with L chirality at the α carbon (**Figure 1.1.1**). The prefix L stands for levo and means that the related compound L-glyceraldehyde rotates polarized light to the left.

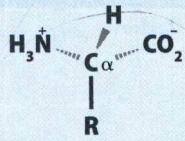


FIGURE 1.1.1
The Cα carbon of amino acids is chiral. This figure shows an L-amino acid.

A D-amino acid (**Figure 1.1.2**) has the opposite **chirality (*1.2)**: D-glyceraldehyde rotates polarized light to the right (dextro).

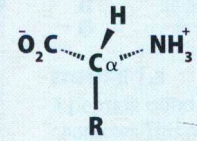


FIGURE 1.1.2
A D-amino acid.

*1.2 Chirality

Any molecule whose reflection in a mirror cannot be superimposed is asymmetric or *chiral*. The two mirror images are called *enantiomers*, or more generally but less specifically *isomers*. Their physical and chemical properties are identical, except that one rotates plane-polarized light to the left and the other rotates it to the right. The most common origin of chirality is carbon atoms that have four nonidentical groups attached to them: for example Cα carbons in amino acids (see Figure 1.1.1). [The exception is glycine, which is not chiral because the Cα has two hydrogens attached and is therefore symmetrical.] The two enantiomers are called L and D. The formal definition of L is as follows: view the Cα with the Hα toward you. If C=O, side chain, N go in a clockwise direction, the amino acid is L, whereas if they are anticlockwise it is D. This nomenclature is related to the organic chemistry (Cahn–Ingold–Prelog) definitions of *S* and *R*: all L-amino acids except cystine are also *S*.

The 20 amino acids are conveniently divided into groups. Four (Asp, Glu, Arg, and Lys) carry a charge at neutral pH: two are positive (basic: Arg and Lys) and two negative (acidic: Asp and Glu). Seven are **hydrophobic** (eight if we include glycine), and the remaining eight have polar groups. Of these, histidine is noteworthy because its pK_a is close to 7. Therefore in a protein at neutral pH it can be either protonated or not, depending on its local environment. Cysteine is also "special" because the side chain is easily oxidized to form the S–S disulfide form, where it is known as cystine. In an extracellular environment, including in the blood, cysteine is usually oxidized to cystine. However, the intracellular environment is normally sufficiently reducing that the dominant form is cysteine. Therefore one commonly finds extracellular proteins that are stabilized by disulfide bridges, whereas disulfides are not usually found in intracellular proteins. (In intracellular proteins, a similar stabilizing role is played by zinc, which binds to a combination of four cysteine or histidine side chains, forming a variety of "zinc finger" structures.) Cysteine also has a fairly low pK_a, making it a good nucleophile (*5.7). It is therefore often found in enzyme active sites.

*1.3 Backbone

This is generally taken to mean the N, Cα and carbonyl CO groups in a protein (**Figure 1.3.1**).

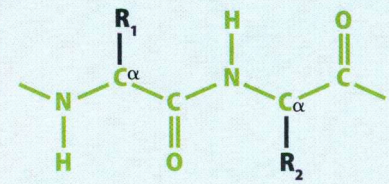


FIGURE 1.3.1
A protein backbone (green).

*1.4 Side chain

The side chain is those parts of a protein that are not the backbone (**Figure 1.4.1**). Each amino acid except glycine has a side chain (Table 1.1).

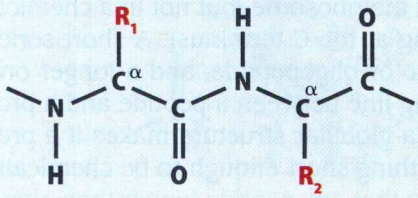


FIGURE 1.4.1
Protein side chains.