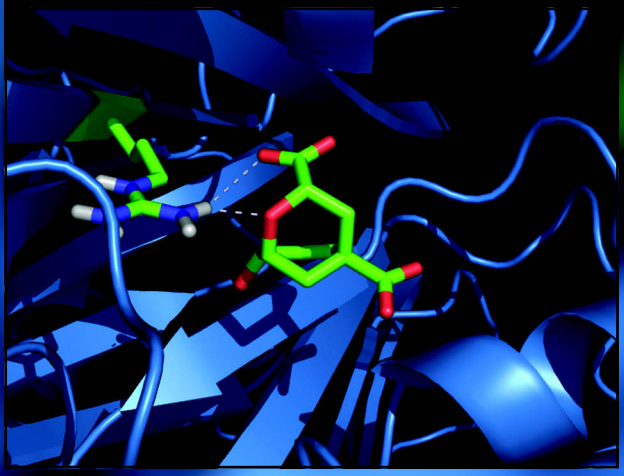
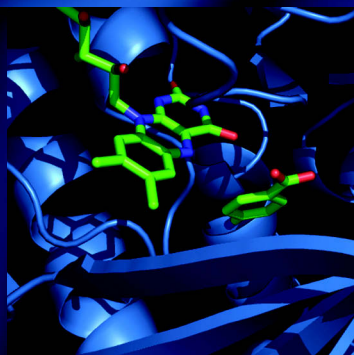
Third Edition

Introduction to Enzyme and Coenzyme Chemistry

T. D. H. Bugg



WILEY



Introduction to Enzyme and Coenzyme Chemistry

Introduction to Enzyme and Coenzyme Chemistry

Third Edition

T. D. H. BUGG
Department of Chemistry, University of Warwick, UK

WILEY
A John Wiley & Sons, Ltd., Publication

This edition first published 2012
© 2012 John Wiley & Sons, Ltd

Registered office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com.

The right of the author to be identified as the author of this work has been asserted in accordance with the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by the UK Copyright, Designs and Patents Act 1988, without the prior permission of the publisher.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The publisher is not associated with any product or vendor mentioned in this book. This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

The publisher and the author make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation any implied warranties of fitness for a particular purpose. This work is sold with the understanding that the publisher is not engaged in rendering professional services. The advice and strategies contained herein may not be suitable for every situation. In view of ongoing research, equipment modifications, changes in governmental regulations, and the constant flow of information relating to the use of experimental reagents, equipment, and devices, the reader is urged to review and evaluate the information provided in the package insert or instructions for each chemical, piece of equipment, reagent, or device for, among other things, any changes in the instructions or indication of usage and for added warnings and precautions. The fact that an organization or Website is referred to in this work as a citation and/or a potential source of further information does not mean that the author or the publisher endorses the information the organization or Website may provide or recommendations it may make. Further, readers should be aware that Internet Websites listed in this work may have changed or disappeared between when this work was written and when it is read. No warranty may be created or extended by any promotional statements for this work. Neither the publisher nor the author shall be liable for any damages arising herefrom.

Library of Congress Cataloging-in-Publication Data

Bugg, T. D. H.
Introduction to enzyme and coenzyme chemistry / T. D. H. Bugg. – 3rd ed.
p. cm.
Includes bibliographical references and index.
ISBN 978-1-119-99595-1 (cloth) – ISBN 978-1-119-99594-4 (pbk.) 1. Enzymes. 2. Coenzymes. I. Title.
QP601.B955 2012
612′.0151–dc23 2012008853

A catalogue record for this book is available from the British Library.

Cloth ISBN: 9781119995951
Paper ISBN: 9781119995944

Typeset in 10/12pt Times by Aptara Inc., New Delhi, India

Instructors can access PowerPoint files of the illustrations presented within this text, for teaching, at: http://booksupport.wiley.com.

Contents

Preface

When I was approached about a 3rd edition of the book, I wanted to enhance the interplay of enzyme active site structure with their catalytic mechanisms and also update the book with some more recent literature examples and topics of current research interest in enzymology. In the 3rd edition I have redrawn the figures showing protein structures using PyMOL software (see Appendix 3). One advantage of PyMOL is that it is easy to prepare high-resolution images, so hopefully the images in the 3rd edition will be a bit sharper than in the 2nd edition. I have used a similar set of examples, but have also added several new examples to supplement the text.

I have also added some new text: in Chapter 3 I have added some new examples of transition state stabilisation and nucleophilic catalysis to Sections 3.4 and 3.6 and added a new Section 3.8 "Desolvation of substrate and active site nucleophiles", using the example of *S. cattleya* fluorinase discovered by the group of Professor David O'Hagan. I've also added some new text and references on the role of protein dynamics in enzyme catalysis, which has been a topic of much discussion and debate in recent years. In Chapter 4 I have mentioned the link between hydrogen tunnelling and temperature-independent kinetic isotope effects, also a topic of current interest in enzymology. In Chapter 5 I have mentioned the discovery of a covalent intermediate in the lysozyme reaction by the group of Professor Stephen Withers, and expanded the discussion of glycosyltransferases. In Chapter 11 I have included a new section 11.6 on "Radical chemistry in DNA repair enzymes", and included a new figure of the *Drosophila melanogaster* (6-4) photolyase; I'd like to thank Professor Thomas Carell for helpful discussions and information on this enzyme. Finally, in Chapter 12 I have included a figure and discussion about the most ancient catalytic reaction of all: the peptidyl transfer reaction on the ribosome. A great deal of new structural data regarding this reaction has emerged in the last 10 years, although the precise catalytic mechanism is still under debate!

I'd like to thank colleagues, researchers and students at Warwick and elsewhere for their support and encouragement.

T. D. H. Bugg
University of Warwick
January 2012

Representation of Protein Three-Dimensional Structures

In the 3rd edition I have used the program PyMOL to draw representations of protein three-dimensional structures. PyMOL was developed by Warren Lyford DeLano and commercialised by DeLano Scientific LLC. The software is freely available to educational users from the WWW (http://www.pymol.org), can be downloaded with instructions for its use and is supported by a helpful Wiki page (see below). There are several packages available for representation of protein structures (e.g. RASMOL, SwissPDB Viewer) that are freely available and straightforward to use. PyMOL allows the user to save all of the information in the current session, to go back and modify later on and to easily render the protein structure images into high-resolution pictures.

In order to view a protein structure, you must first download the PDB file from the Brookhaven Protein Database, which contains all the data for published X-ray crystal structures and NMR structures of proteins and nucleic acids. I have included the PDB filename for each of the pictures that I have drawn in the figure legend. I recommend that you download a few of these and try viewing each one on a computer screen; you can turn the structure around and get a really good feel for the three-dimensional structure of the protein. You can download the PDB file from http://www.rcsb.org/pdb, or several other web sites.

Once you have downloaded the PDB file (in PDB text format), then you run the PyMOL program and open the PDB structure file to view the structure. You can view the protein backbone in several different ways: as individual atoms (lines or sticks) or protein secondary structure (ribbon or cartoon). In most of the pictures in this edition I have drawn the protein backbone in cartoon format. I have then selected certain catalytic amino acid residues and highlighted them in red, and selected any bound substrate analogues or coenzymes and highlighted them in black or red. In preparing figures for the book I used only black and red, but on a computer screen you can use a wide range of colours and prepare your own multi-colour pictures!

References

W.L. DeLano, “The PyMOL Molecular Graphics System”, DeLano Scientific, San Carlos, CA (2002).
PyMOL home page: http://www.pymol.org
PyMOL Wiki page: http://www.pymolwiki.org

1

From Jack Beans to Designer Genes

1.1 Introduction

Enzymes are giant macromolecules which catalyse biochemical reactions. They are remarkable in many ways. Their three-dimensional structures are highly complex, yet they are formed by spontaneous folding of a linear polypeptide chain. Their catalytic properties are far more impressive than synthetic catalysts which operate under more extreme conditions. Each enzyme catalyses a single chemical reaction on a particular chemical substrate with very high enantioselectivity and enantiospecificity at rates which approach "catalytic perfection". Living cells are capable of carrying out a huge repertoire of enzyme-catalysed chemical reactions, some of which have little or no precedent in organic chemistry. In this book I shall seek to explain from the perspective of organic chemistry what enzymes are, how they work, and how they catalyse many of the major classes of enzymatic reactions.

1.2 The discovery of enzymes

Historically, biological catalysis has been used by mankind for thousands of years, ever since fermentation was discovered as a process for brewing and bread-making in ancient Egypt. It was not until the 19th century A.D. however that scientists addressed the question of whether the entity responsible for processes such as fermentation was a living species or a chemical substance. In 1897 Eduard Buchner published the observation that cell-free extracts of yeast containing no living cells were able to carry out the fermentation of sugar to alcohol and carbon dioxide. He proposed that a species called "zymase" found in yeast cells was responsible for fermentation. The biochemical pathway involved in fermentation was subsequently elucidated by Embden and Meyerhof – the first pathway to be discovered.

The exquisite selectivity of enzyme catalysis was recognised as early as 1894 by Emil Fischer, who demonstrated that the enzyme which hydrolyses sucrose, which he called "invertin", acts only upon α-D-glucosides, whereas a second enzyme "emulsin" acts only upon β-D-glucosides. He deduced that these two enzymes must consist of "asymmetrically built

Introduction to Enzyme and Coenzyme Chemistry, Third Edition. T. D. H. Bugg.
© 2012 John Wiley & Sons, Ltd. Published 2012 by John Wiley & Sons, Ltd.

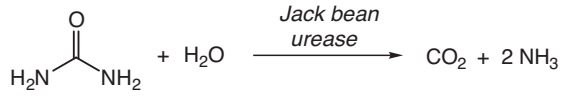


Figure 1.1 *Urease.*

molecules", and that "the enzyme and glucoside must fit each other like a lock and key to be able to exert a chemical influence upon each other". Fischer's "lock and key" hypothesis remained a powerful metaphor of enzyme action for many years. The crystallisation in 1926 of the enzyme urease (see Figure 1.1) from Jack beans by Sumner proved beyond doubt that biological catalysis was carried out by a chemical substance.

The recognition that biological catalysis is mediated by enzymes heralded the growth of biochemistry as a subject, and the elucidation of the metabolic pathways catalysed by enzymes. Each reaction taking place on a biochemical pathway is catalysed by a specific enzyme. Without enzyme catalysis the uncatalysed chemical process would be too slow to sustain life. Enzymes catalyse reactions involved in all facets of cellular life: metabolism (the production of cellular building blocks and energy from food sources); biosynthesis (how cells make new molecules); detoxification (the breakdown of toxic foreign chemicals); and information storage (the processing of deoxyribonucleic acids).

In any given cell there are present several thousand different enzymes, each catalysing its specific reaction. How does a cell know when it needs a particular enzyme? The production of enzymes, as we shall see in Chapter 2, is controlled by a cell's DNA, both in terms of the specific structure of a particular enzyme and the amount which is produced. Thus different cells in the same organism have the ability to produce different types of enzymes and to produce them in differing amounts according to the cell's requirements.

Since the majority of the biochemical reactions involved in cellular life are common to all organisms, a given enzyme will usually be found in many or all organisms, albeit in different forms and amounts. By looking closely at the structures of enzymes from different organisms which catalyse the same reaction, we can in many cases see similarities between them. These similarities are due to the evolution and differentiation of species by natural selection. So by examining closely the similarities and differences of an enzyme from different species we can trace the course of molecular evolution, as well as learning about the structure and function of the enzyme itself.

Recent developments in biochemistry, molecular biology and X-ray crystallography now allow a far more detailed insight into how enzymes work at a molecular level. We now have the ability to determine the amino acid sequence of enzymes with relative ease, whilst the technology for solving the three-dimensional structure of enzymes is developing apace. We also have the ability to analyse their three-dimensional structures using molecular modelling and then to change the enzyme structure rationally using site-directed mutagenesis. We are now starting to enter the realms of enzyme engineering, where by rational design we can modify the genes encoding specific enzymes, creating the "designer genes" of the title. These modified enzymes could in future perhaps be used to catalyse new types of chemical reactions, or via gene therapy to correct genetic defects in cellular enzymes which would otherwise lead to human diseases.

1.3 The discovery of coenzymes

At the same time as the discovery of enzymes in the late 19th and early 20th centuries, a class of biologically important small molecules was being discovered which had remarkable properties to cure certain dietary disorders. These molecules were christened the vitamins, a corruption of the phrase "vital amines" used to describe their dietary importance (several of the first-discovered vitamins were amines, but this is not true of all the vitamins). The vitamins were later found to have very important cellular roles, shown in Table 1.1.

The first demonstration of the importance of vitamins in the human diet took place in 1753. A Scottish naval physician James Lind found that the disease scurvy, prevalent amongst mariners at that time, could be avoided by deliberately including green vegetables or citrus fruits in the sailors' diets. This discovery was used by Captain James Cook to maintain the good health of his crew during his voyages of exploration in 1768–1776. The active ingredient was elucidated much later as vitamin C, ascorbic acid.

The first vitamin to be identified as a chemical substance was thiamine, lack of which causes the limb paralysis beriberi. This nutritional deficiency was first identified in the Japanese Navy in the late 19th century. The incidence of beriberi in sailors was connected with their diet of polished rice by Admiral Takaki, who eliminated the ailment in 1885 by improvement of the sailors' diet. Subsequent investigations by Eijkman identified a substance present in rice husks able to cure beriberi. This vitamin was subsequently shown to be an essential "cofactor" in cellular decarboxylation reactions, as we shall see in Chapter 7.

Over a number of years the family of vitamins shown in Table 1.1 was identified and their chemical structures elucidated. Some like vitamin C have simple structures, whilst others like vitamin B12 have very complex structures. It has taken much longer to elucidate the molecular details of their biochemical mode of action. Many of the vitamins are in fact co-enzymes: small organic co-factors which are used by certain types of enzyme in order to carry out particular classes of reaction. Table 1.1 gives a list of the co-enzymes that we are going to encounter in this book.

1.4 The commercial importance of enzymes in biosynthesis and biotechnology

Many plants and micro-organisms contain natural products which possess potent biological activities. The isolation of these natural products has led to the discovery of many biologically active compounds such as quinine, morphine, and penicillin (see Figure 1.2) which have been fundamental to the development of modern medicine.

The process of natural product discovery continues today, with the recent identification of important compounds such as cyclosporin A, a potent immunosuppressant which has dramatically reduced the rejection rate in organ transplant operations; and taxol, an extremely potent anticancer drug isolated from yew bark (see Figure 1.3).

Many of these natural products are structurally so complex that it is not feasible to synthesise them in the laboratory at an affordable price. Nature, however, is able to bio-synthesise these molecules with apparent ease using enzyme-catalysed biosynthetic pathways. Hence there is considerable interest in elucidating the biosynthetic pathways for important natural products

Table 1.1 *The vitamins.*

Vitamin	Chemical name	Deficiency disease	Biochemical function	Coenzyme chemistry
A	Retinol	Night blindness	Visual pigments	–
B_1	Thiamine	Beriberi	Coenzyme (TPP)	Decarboxylation of α-keto acids
B_2	Riboflavin	Skin lesions	Coenzyme (FAD, FMN)	$1e^-/2e^-$ redox chemistry
Niacin	Nicotinamide	Pellagra	Coenzyme (NAD)	Redox chemistry
B_6	Pyridoxal	Convulsions	Coenzyme (PLP)	Reactions of α-amino acids
B_{12}	Cobalamine	Pernicious anaemia	Coenzyme	Radical re-arrangements
C	Ascorbic acid	Scurvy	Coenzyme, anti-oxidant	Redox agent (collagen formation)
D	Calciferols	Rickets	Calcium homeostasis	–
E	Tocopherols	Newborn haemolytic aneamia	Anti-oxidant	–
H	Biotin	Skin lesions	Coenzyme	Carboxylation
K	Phylloquinone	Bleeding disorders	Coenzyme, anti-oxidant	Carboxylation of glutamyl peptides
	Folic acid	Megaloblastic anaemia	Coenzyme (tetrahydrofolate)	1-carbon transfers
	Pantothenic acid	Burning foot syndrome	Coenzyme (CoA, phosphopantotheine) Acyl transfer	

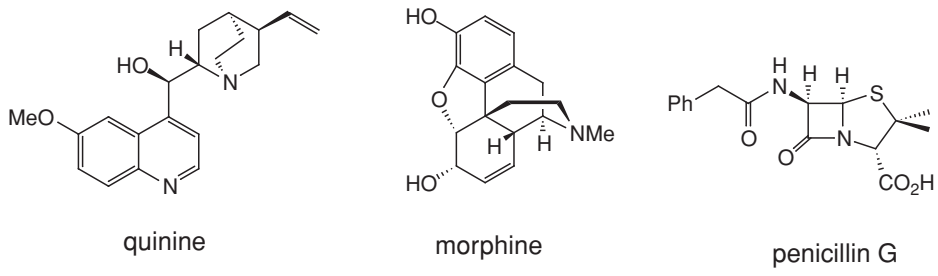


Figure 1.2 *Structures of quinine, morphine, penicillin.*

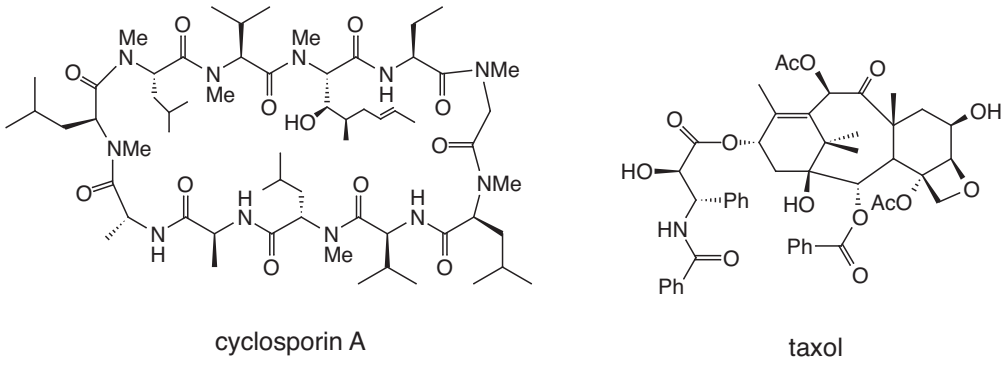


Figure 1.3 *Structures of CsA, taxol.*

and using the enzymes to produce natural products *in vitro*. One example of this is the industrial production of semi-synthetic penicillins using a naturally occurring enzyme penicillin acylase (see Figure 1.4). Penicillin G which is obtained from growing *Penicillium* mould has certain clinical disadvantages; enzymatic deacylation and chemical re-acylation give a whole range of "semi-synthetic" penicillins which are clinically more useful.

The use of enzyme catalysis for commercial applications is an exciting area of the biotechnology industry. One important application that we shall encounter is the use of enzymes in asymmetric organic synthesis. Since enzymes are highly efficient catalysts that work under mild conditions and are enantiospecific, they can in many cases be used on a practical scale to resolve racemic mixtures of chemicals into their optically active components. This is becoming increasingly important in drug synthesis, since one enantiomer of a drug usually

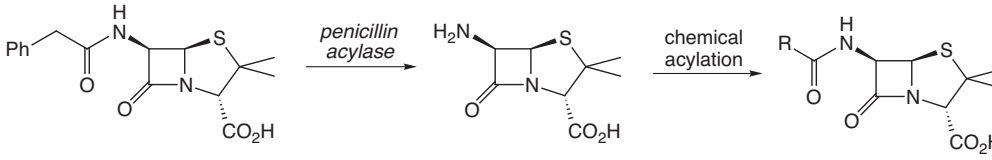


Figure 1.4 *Penicillin acylase.*

Table 1.2 *Commercial applications of enzyme inhibitors.*

Antibacterial agents	**Penicillins** and **cephalosporins** inactivate the *transpeptidase* enzyme which normally makes cross-links in the bacterial cell wall (peptidoglycan), leading to weakened cell walls and eventual cell lysis. **Streptomycin** and **kanamycin** inhibit protein synthesis on bacterial ribosomes, whereas mammalian ribosomes are less affected.
Antifungal agents	**Ketoconazole** inhibits *lanosterol 14α-demethylase*, an enzyme involved in the biosynthesis of an essential steroid component of fungal cell membranes. **Nikkomycin** inhibits *chitin synthase*, an enzyme involved in making the chitin cell walls of fungi.
Antiviral agents	**AZT** inhibits the *reverse transcriptase* enzyme required by the HIV virus in order to replicate its own DNA.
Insecticides	Organophosphorus compounds such as **dimethoate** derive their lethal activity from the inhibition of the insect enzyme *acetylcholinesterase* involved in the transmission of nerve impulses.
Herbicides	**Glyphosate** inhibits the enzyme *EPSP synthase* which is involved in the biosynthesis of the essential amino acids phenylalanine, tyrosine and tryptophan (see Chapter 8.5).

has very different biological properties from the other. The unwanted enantiomer might have detrimental side-effects, as in the case of thalidomide, where one enantiomer of the drug was useful in relieving morning sickness in pregnant women, but the other enantiomer caused serious deformities in the newborn child when the racemic drug was administered.

1.5 The importance of enzymes as targets for drug discovery

If there is an *essential* enzyme found uniquely in a certain class of organisms or cell type, then a selective *inhibitor* of that enzyme could be used for selective toxicity against that organism or cell type. Similarly, if there is a significant difference between a particular enzyme found in bacteria as compared with the same enzyme in humans, then a selective inhibitor could be developed for the bacterial enzyme. If this inhibitor did not inhibit the human enzyme, then it could be used as an antibacterial agent. Thus, *enzyme inhibition is a basis for drug discovery*.

This principle has been used for the development of a range of pharmaceutical and agrochemical agents (Table 1.2) – we shall see examples of important enzyme targets later in the book. In many cases resistance to these agents has emerged due to mutation in the structures of the enzyme targets. This has provided a further incentive to study the three-dimensional structures of enzyme targets, and has led to the development of powerful molecular modelling software for analysis of enzyme structure and *de novo design* of enzyme inhibitors.

The next two chapters are "theory" chapters on enzyme structure and enzyme catalysis, followed by a "practical" chapter on methods used to study enzymatic reactions. Chapters 5–11 cover each of the major classes of enzymatic reactions, noting each of the coenzymes used for enzymatic reactions. Finally there is a brief introduction in Chapter 12 to other types of biological catalysis. In cases where discussion is brief the interested reader will find references to further reading at the end of each chapter.

2

All Enzymes Are Proteins

2.1 Introduction

Enzymes are giant molecules. Their molecular weight varies from 5,000 to 5,000,000 Da, with typical values in the range 20,000–100,000 Da. At first sight this size suggests a bewildering complexity of structure, yet we shall see that enzymes are structurally assembled in a small number of steps in a fairly simple way.

Enzymes belong to a larger biochemical family of macromolecules known as proteins. The common feature of proteins is that they are polypeptides: their structure is made up of a linear sequence of α-amino acid building blocks joined together by amide linkages. This linear polypeptide chain then "folds" to give a unique three-dimensional structure.

2.2 The structures of the L-α-amino acids

Proteins are composed of a family of 20 α-amino acid structural units whose general structure is shown in Figure 2.1. α-Amino acids are chiral molecules: that is, their mirror image is not superimposable upon the original molecule.

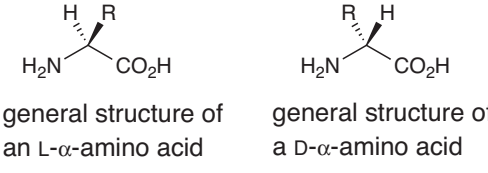


Figure 2.1 *General structure of L- & D-amino acids.*

Each α-amino acid can be found as either the L- or D-isomer depending on the configuration at the α-carbon atom (except for glycine where R=H). All proteins are composed only of L-amino acids, consequently enzymes are inherently chiral molecules – an important point.

Introduction to Enzyme and Coenzyme Chemistry, Third Edition. T. D. H. Bugg.
© 2012 John Wiley & Sons, Ltd. Published 2012 by John Wiley & Sons, Ltd.

D-amino acids are rare in biological systems, although they are found in a number of natural products and notably in the peptidoglycan layer of bacterial cell walls (see Chapter 9).

The α-amino acids used to make up proteins number only twenty, whose structures are shown in Figure 2.2. The differences between these twenty lie in the nature of the sidechain R. The simplest amino acids are glycine (abbreviated Gly or simply G), which has no sidechain,

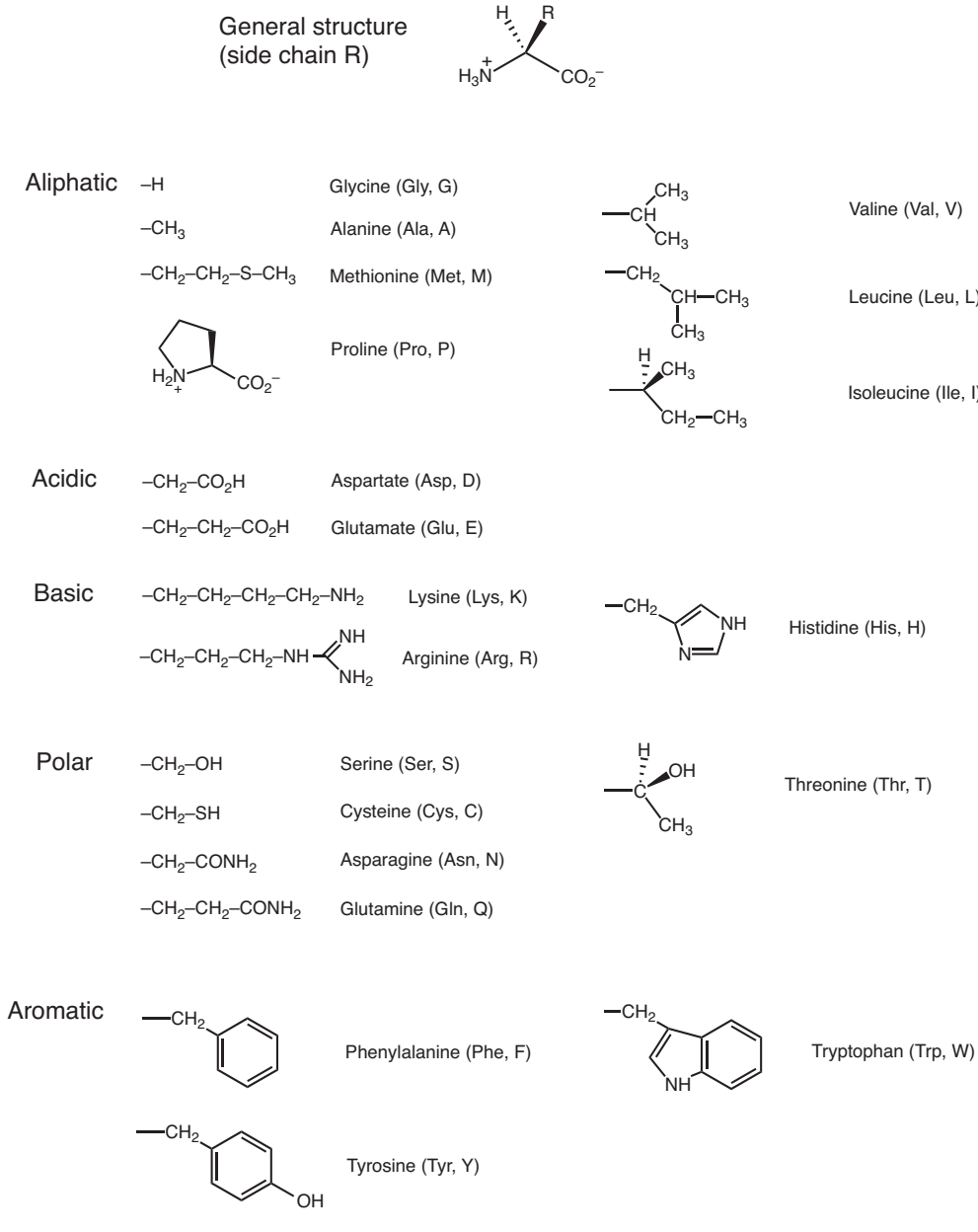


Figure 2.2 *The side-chains of the 20 L-amino acids found in proteins. Whole amino acid structure shown for proline.*

and alanine (Ala or A), whose sidechain is a methyl group. A number of sidechains are hydrophobic ("water-hating") in character, for example the thioether of methionine (Met); the branched aliphatic sidechains of valine (Val), leucine (Leu) and isoleucine (Ile); and the aromatic sidechains of phenylalanine (Phe) and tryptophan (Trp). The remainder of the amino acid sidechains are hydrophilic ("water-loving") in character. Aspartic acid (Asp) and glutamic acid (Glu) contain carboxylic acid sidechains, and their corresponding primary amides are found as asparagine (Asn) and glutamine (Gln). There are three basic sidechains consisting of the ε-amino group of lysine (Lys), the guanidine group of arginine (Arg), and the imidazole ring of histidine (His). The polar nucleophilic sidechains that will assume a key role in enzyme catalysis are the primary hydroxyl of serine (Ser), the secondary hydroxyl of threonine (Thr), the phenolic hydroxyl group of tyrosine (Tyr), and the thiol group of cysteine (Cys).

The nature of the sidechain confers certain physical and chemical properties upon the corresponding amino acid, and upon the polypeptide chain in which it is located. The amino acid sidechains are therefore of considerable structural importance, and as we shall see in Chapter 3 they play key roles in the catalytic function of enzymes.

2.3 The primary structure of polypeptides

To form the polypeptide chain found in proteins each amino acid is linked to the next via an amide bond, forming a linear sequence of 100–1,000 amino acids – this is the primary structure of the protein. A portion of the amino-terminal (or N-terminal) end of a polypeptide is shown in Figure 2.3, together with the abbreviated representations for this peptide sequence.

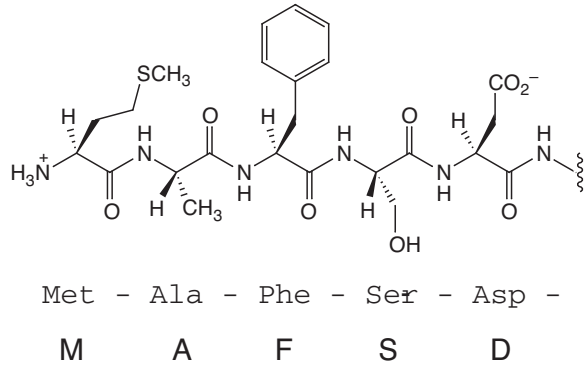


Figure 2.3 *Structure of the N-terminal portion of linear polypeptide chain.*

The sequence of amino acids in the polypeptide chain is all-important. It contains all the information to confer both the three-dimensional structure of proteins in general and the catalytic activity of enzymes in particular. How is this amino acid sequence controlled? It is specified by the nucleotide sequence of the corresponding *gene*, the piece of DNA (deoxyribonucleic acid) which encodes for that particular protein in that particular organism. To give an idea of how this is achieved, I will give a simplified account of how the polypeptide

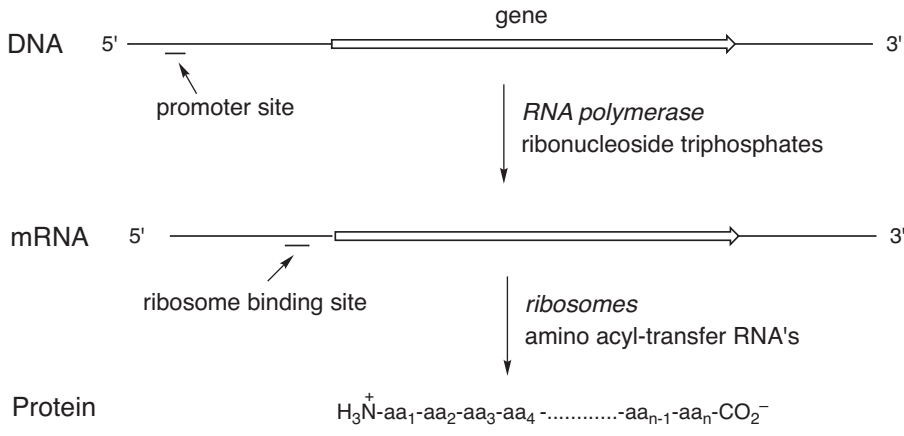


Figure 2.4 *Pathway for protein biosynthesis from corresponding gene, via messenger RNA.*

sequence is derived from the gene sequence. For a more detailed description the reader is referred to biochemical textbooks.

Genes are composed of four deoxyribo-nucleotides (or "bases"): deoxyadenine (dA), deoxycytidine (dC), deoxyguanine (dG) and thymidine (dT), arranged in a specific linear sequence. To give some idea of size, a typical gene might consist of a sequence of 1,000 nucleotide bases encoding the information for the synthesis of a protein of approximately 330-amino acids, whose molecular weight would be 35–40 kDa.

How is the sequence encoded? First the deoxyribo-nucleotide sequence of the DNA strand is transcribed into messenger RNA containing the corresponding ribo-nucleotides adenine (A), cytidine (C), guanine (G) and uridine (U, corresponding to dT). The RNA strand is then translated into protein by the biosynthetic machinery known as ribosomes, as shown in Figure 2.4. The RNA sequence is translated into protein in sets of three nucleotide bases, one set of three bases being known as a "triplet codon". Each codon encodes a single amino acid. The code defining which amino acid is derived from which triplet codon is the "universal genetic code", shown in Figure 2.5. This universal code is followed by the protein biosynthetic machinery of all organisms.

As an example we shall consider in Figure 2.6 the N-terminal peptide sequence Met-Ala-Phe-Ser-Asp illustrated in Figure 2.3. The first amino acid at the N-terminus of each protein is always methionine, whose triplet codon is AUG. The next triplet GCC encodes alanine; UUC encodes phenylalanine; UCC encodes serine; and GAC encodes aspartate. Translation then continues in triplets until one of three "stop" codons is reached; at this point protein translation stops. Note that for most amino acids there is more than one possible codon: thus if UUC is changed to UUU, phenylalanine is still encoded, but if changed to UCC then serine is encoded as above.

In this way the nucleotide sequence of the gene is translated into the amino acid sequence of the encoded protein. An important practical consequence is that the amino acid sequence of an enzyme can be determined by nucleotide sequencing of the corresponding gene, which is now the most convenient way to determine a protein sequence.

AAA	Lys	ACA	Thr	AGA	Arg	AUA	Ile
AAG	Lys	ACG	Thr	AGG	Arg	AUG	Met
AAC	Asn	ACC	Thr	AGC	Ser	AUC	Ile
AAU	Asn	ACU	Thr	AGU	Ser	AUU	Ile
CAA	Gln	CCA	Pro	CGA	Arg	CUA	Leu
CAG	Gln	CCG	Pro	CGG	Arg	CUG	Leu
CAC	His	CCC	Pro	CGC	Arg	CUC	Leu
CAU	His	CCU	Pro	CGU	Arg	CUU	Leu
GAA	Glu	GCA	Ala	GGA	Gly	GUA	Val
GAG	Glu	GCG	Ala	GGG	Gly	GUG	Val
GAC	Asp	GCC	Ala	GGC	Gly	GUC	Val
GAU	Asp	GCU	Ala	GGU	Gly	GUU	Val
UAA	Stop	UCA	Ser	UGA	Stop	UUA	Leu
UAG	Stop	UCG	Ser	UGG	Trp	UUG	Leu
UAC	Tyr	UCC	Ser	UGC	Cys	UUC	Phe
UAU	Tyr	UCU	Ser	UGU	Cys	UUU	Phe

Figure 2.5 *The universal genetic code.*

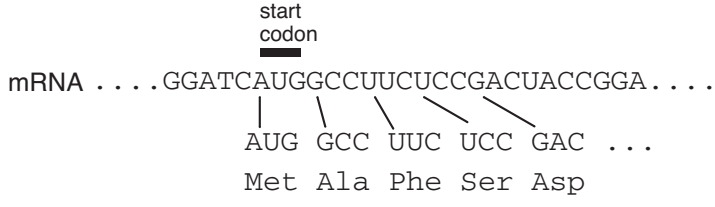


Figure 2.6 *Translation of mRNA to protein.*

2.4 Alignment of amino acid sequences

Most biochemical reactions are found in more than one organism, in some cases in all living cells. If the enzymes which catalyse the same reaction in different organisms are purified and their amino acid sequences are determined, then we often see similarity between the two sequences. The degree of similarity is usually highest in enzymes from organisms which have evolved recently on an evolutionary timescale. The implication of such an observation is that the two enzymes have evolved divergently from a common ancestor.

Over a long period of time, changes in the DNA sequence of a gene can occur by random mutation or by several types of rare mistakes in DNA replication. Many of these mutations will lead to a change in the encoded protein sequence in such a way that the mutant protein is inactive. These mutations are likely to be lethal to the cell and are hence not passed down to the next generation. However, mutations which result in minor modifications to non-essential residues in an enzyme will have little effect on the activity of the enzyme, and will therefore be passed on to the next generation.

```
Alignment of N-terminal 15 amino acids of four sequences in 3-letter codes:

                              1               5                   10                  15
E. coli MhpB                   Met His Ala Tyr Leu His Cys Leu Ser His Ser Pro Leu Val Gly
A. eutrophus MpcI              Met Pro Ile Gln Leu Glu Cys Leu Ser His Thr Pro Leu His Gly
P. paucimobilis LigB   Met Ala Arg Val Thr Thr Gly Ile Thr Ser Ser His Ile Pro Ala Leu Gly
E. coli HpcB           Met Gly Lys Leu Ala Leu Ala Ala Lys Ile Thr His Val Pro Ser Met Tyr
                                                           +   *       *

Alignment of N-terminal 60 amino acids of two sequences in 1-letter codes:

                    1          11         21         31         41         51
E. coli MhpB        MHAYLHCLSH SPLVGYVDPA QEVLDEVNGV IASARERIAA FSPELVVLFA PDHYNGFFYD
A. eutrophus MpcI   MPIQLECLSH TPLHGYVDPA PEVVAEVERV QAAARDRVRA FDPELVVVFA PDHFNGFFYD
                     *    **** +** ******  **+ **  *  *+***+*+ * * ******+** ***+*******

* = identically conserved residue  + = functionally conserved residue
```

Figure 2.7 *Amino acid sequence alignment.*

So if we look at an alignment of amino acid sequences of "related" enzymes from different organisms, we would expect that catalytically important amino acid residues would be invariant or "conserved" in all species. In this way sequence alignments can provide clues for identifying important amino acid residues in the absence of an X-ray crystal structure. For example, in Figure 2.7 there is an alignment of the N-terminal portion of the amino acid sequence of a dioxygenase enzyme MhpB from *Escherichia coli* with "related" dioxygenase enzymes from *Alcaligenes eutrophus* (MpcI) and *Pseudomonas* (LigB) and another *E. coli* enzyme HpcB. Clearly there are a small number of conserved residues (indicated by a *) which are very important for activity, and a further set of residues for which similar amino acid sidechains are found (e.g. hydroxyl-containing serine and threonine, indicated with a +).

Furthermore, sequence similarity is sometimes observed between different enzymes which catalyse similar reactions or use the same cofactor, giving rise to "sequence motifs" found in a family of enzymes. We shall meet some examples of sequence motifs later in the book.

2.5 Secondary structures found in proteins

When the linear polypeptide sequence of the protein is formed inside cells by ribosomes, a remarkable thing happens: the polypeptide chain spontaneously folds to form the three-dimensional structure of the protein. All the more remarkable is that from a sequence of 100–1,000 amino acids a *unique* stable three-dimensional structure is formed. It has been calculated that if the protein folding process were to sample each of the available conformations then it would take longer than the entire history of the universe – yet in practice it takes a few seconds! The mystery of protein folding is currently a topic of intense research, and the interested reader is referred to specialist articles on this topic. Factors that seem to be important in the folding process are: 1) packing of hydrophobic amino acid sidechains and exclusion of solvent water; 2) formation of specific non-covalent interactions; 3) formation of secondary structures.

Secondary structure is the term given to local regions (10–20 amino acids) of stable, ordered three-dimensional structures held together by hydrogen-bonding, that is non-covalent bonding between acidic hydrogens (O—H, N—H) and lone pairs as shown in Figure 2.8.

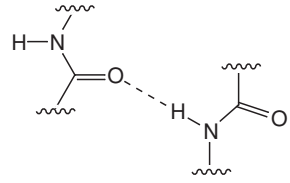


Figure 2.8 *A hydrogen bond.*

There are at least three stable forms of secondary structure commonly observed in proteins: the α-helix, the β-sheet, and the β-turn. The α-helix is a helical structure formed by a single polypeptide chain in which hydrogen-bonds are formed between the carbonyl oxygen of one amide linkage and the N—H of the amide linkage four residues ahead in the chain, as shown in Figure 2.9.

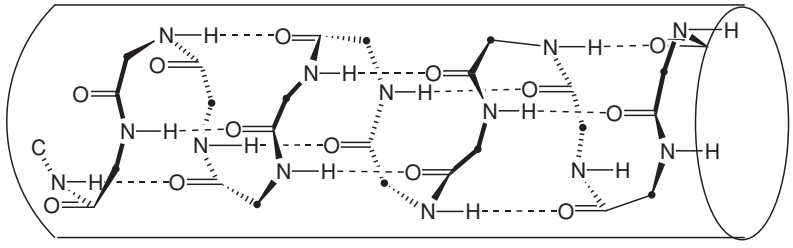


Figure 2.9 *Structure of an α-helix. Positions of amino acid side-chains are indicated with dots.*

In this structure each of the amide linkages forms two specific hydrogen-bonds, making it a very stable structural unit. All of the amino acid sidechains point outwards from the pitch of the helix, consequently amino acid sidechains which are four residues apart in the primary sequence will end up close in space. Interactions between such sidechains can lead to further favourable interactions within the helix, or with other secondary structures. A typical α-helix is shown in Figure 2.10A, showing the positions of the sidechains of the amino acid residues.

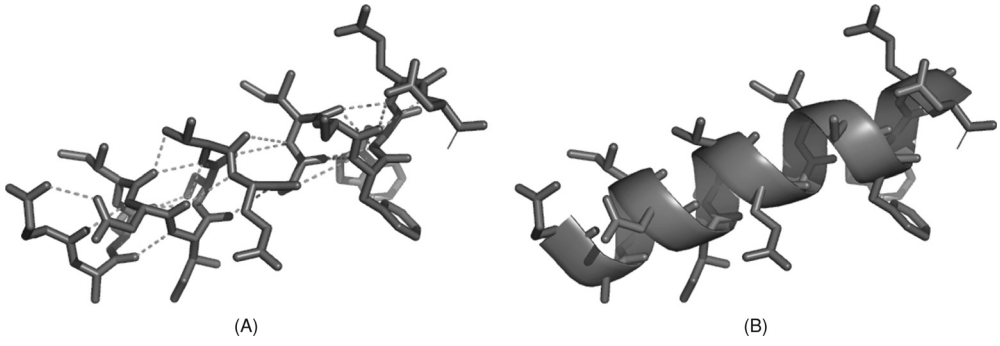


Figure 2.10 *Structure of an α-helix (A) showing amino acid sidechains and hydrogen bonds between amino acid sidechains; (B) showing helix in ribbon form.*

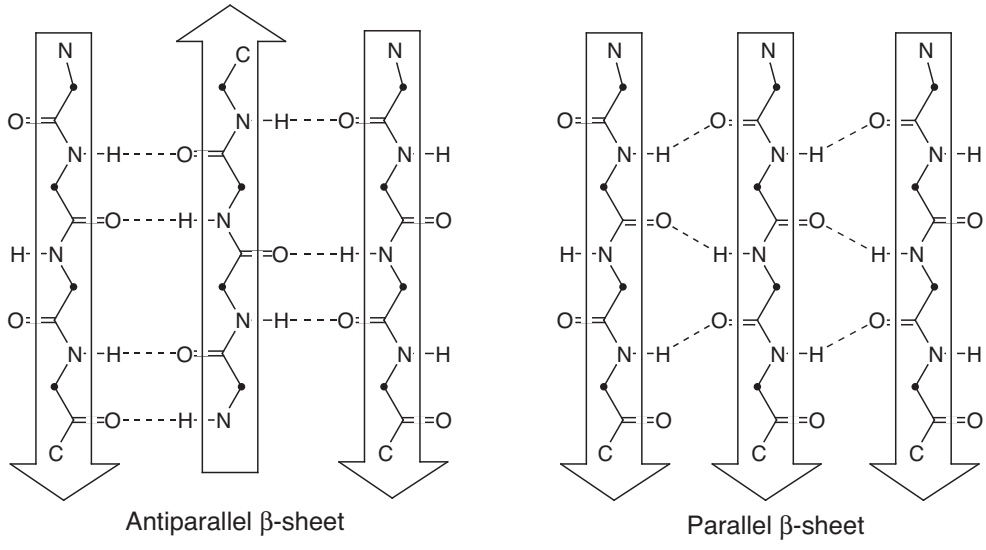


Figure 2.11 *Structures of parallel and antiparallel β-sheets.*

In Figure 2.10B, the same helix is drawn in "ribbon" form, a convenient representation which is used for drawing protein structures.

The β-sheet is a structure formed by two or more linear polypeptide strands, held together by a series of inter-strand hydrogen bonds. There are two types of β-sheet structures: parallel β-sheets, in which the peptide strands both proceed in the same amino-to-carboxyl direction; and antiparallel, in which the peptide strands proceed in opposite directions. Both types are illustrated in Figure 2.11. Figure 2.12A shows an example of two antiparallel β-sheets in a protein structure, with Figure 2.12B showing the same β-sheets in "ribbon" form.

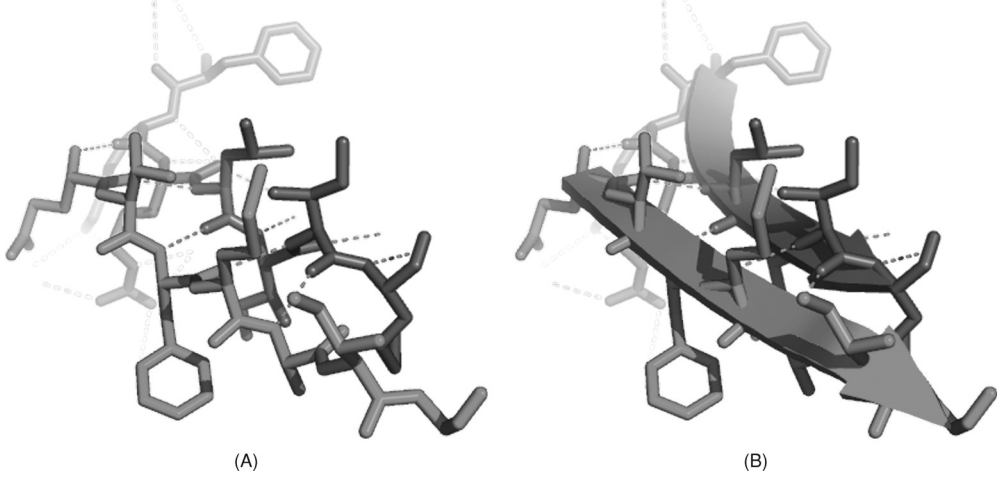


Figure 2.12 *Two antiparallel β-sheets (A) showing amino acid sidechains and hydrogen bonds between sheets; (B) in ribbon format.*

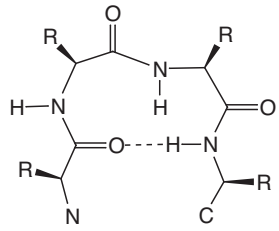


Figure 2.13 *Structure of a β-turn.*

The β-turn is a structure often formed at the end of a β-sheet which leads to a 180° turn in the direction of the peptide chain. An example of a β-turn is shown in Figure 2.13, where the role of hydrogen bonding in stabilising such structures can be seen.

2.6 The folded tertiary structure of proteins

The three-dimensional structure of protein subunits, known as the tertiary structure, arises from packing together elements of secondary structure to form a stable global conformation, which in the case of enzymes is catalytically active. The packing of secondary structural units usually involves burying hydrophobic amino acid sidechains on the inside of the protein and positioning hydrophilic amino acid sidechains on the surface.

Although in theory the number of possible protein tertiary structures is virtually infinite, in practice proteins are often made up of common structural motifs, from which the protein structure can be categorised. Common families of protein structure are: 1) α-helical proteins; 2) α/β structures; 3) antiparallel β structures. Members of each class are illustrated below, with α-helices and β-sheets represented in ribbon form. The α-helical proteins are made up only of α-helices which pack onto one another to form the tertiary structure. Many of the heme-containing cytochromes which act as electron carriers (see Chapter 6) are four-helix "bundles", illustrated in Figure 2.14 in the case of cytochrome b562. The family of globin oxygen carriers, including hemoglobin, consist of a more complex α-helical tertiary structure. The α/β structures consist of regular arrays of β-sheet-α-helix-*parallel*-β-sheet structures. The redox flavoprotein flavodoxin contains five such parallel β-sheets, forming a twisted β-sheet surface interwoven with α-helices, as shown in Figure 2.15. Antiparallel β structures consist of regular arrays of β-sheet-β-turn-*antiparallel* β-sheet. For example, the metallo-enzyme superoxide dismutase contains a small barrel of antiparallel β-sheets, as shown in Figure 2.16.

Frequently proteins consist of a number of "domains", each of which contains a region of secondary structure. Sometimes a particular domain has a specific function, such as binding a substrate or cofactor. Larger proteins often consist of more than one tertiary structure, which fit together to form the active "quaternary" structure. In some cases a number of identical subunits can bind together to form a homodimer (two identical subunits), trimer or tetramer, or in other cases non-identical subunits fit together to form highly complex quaternary structures. One familiar example is the mammalian oxygen transport protein hemoglobin, which consists of a tetramer of identical 16 kDa subunits.

How are protein tertiary structures determined experimentally? The most common method for solving three-dimensional structures of proteins is to use X-ray crystallography, which

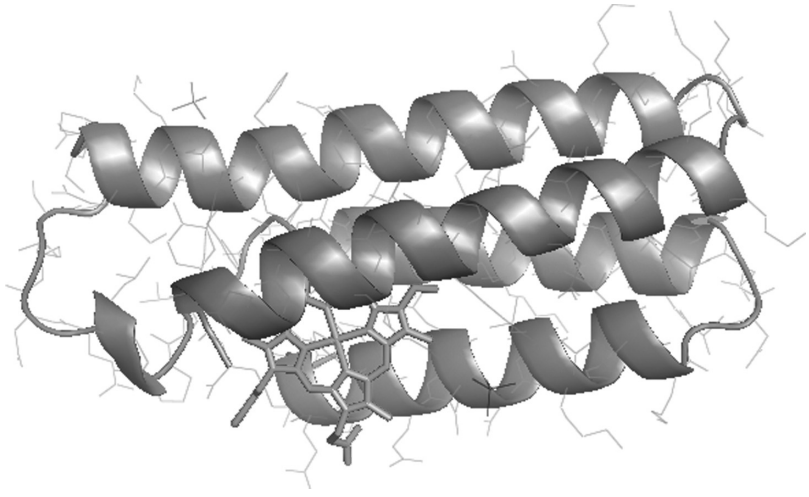


Figure 2.14 *Structure of cytochrome b_{562} (PDB file 256B), a four-helix bundle protein. Heme cofactor shown in red.*

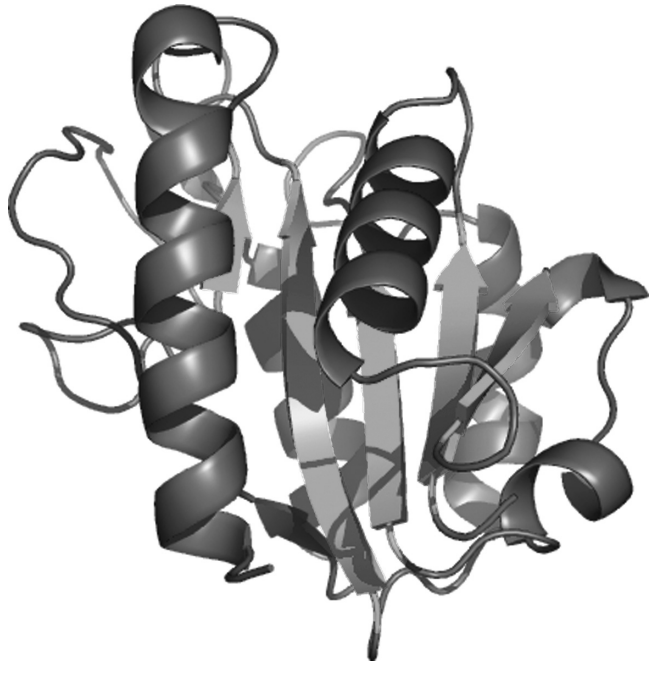


Figure 2.15 *Structure of flavodoxin (PDB file 1AHN), a redox carrier protein containing five parallel β-sheets, each connected by an intervening α-helix. Parallel β-sheets shown in red.*

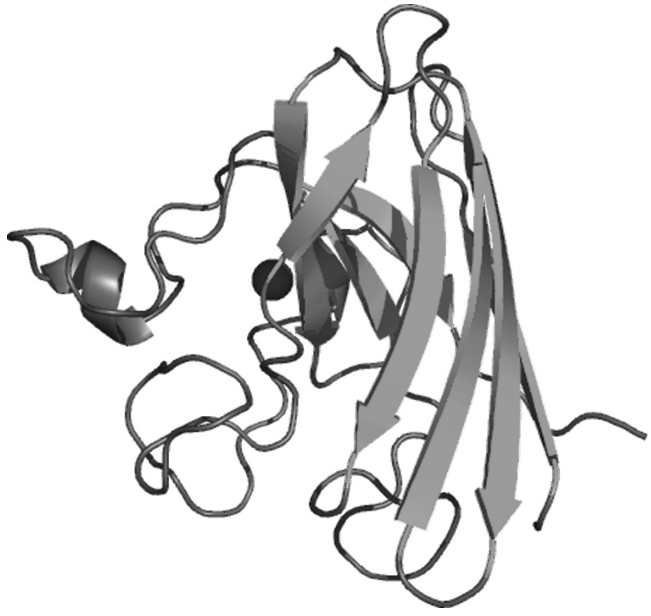


Figure 2.16 *Structure of superoxide dismutase (PDB file 1CB4), a β-barrel protein containing eight antiparallel β-sheets. Antiparallel β-sheets shown in red.*

involves crystallisation of the protein, and analysis of the diffraction pattern obtained from X-ray irradiation of the crystal. The first protein structure to be solved by this method was lysozyme in 1965, since which time several hundred crystal structures have been solved. Recent advances in nuclear magnetic resonance (NMR) spectroscopy have reached the point where the three-dimensional structures of small proteins (<15 kDa) in solution can be solved using multi-dimensional NMR techniques.

2.7 Enzyme structure and function

All enzymes are proteins, but not all proteins are enzymes, the difference being that enzymes possess catalytic activity. The part of the enzyme tertiary structure which is responsible for the catalytic activity is called the "active site" of the enzyme, and often makes up only 10–20% of the total volume of the enzyme. This is where the enzyme chemistry takes place.

The active site is usually a hydrophilic cleft or cavity containing an array of amino acid sidechains which bind the substrate and carry out the enzymatic reaction, as shown in Figure 2.17 (A). In some cases the enzyme active site also binds one or more cofactors which assist in the catalysis of particular types of enzymatic reactions, as shown in Figure 2.17 (B).

How does the enzyme bind the substrate? One of the hallmarks of enzyme catalysis is its high substrate selectivity, which is due to a series of highly specific non-covalent enzyme-substrate binding interactions. Since the active site is chiral, it is naturally able to bind one

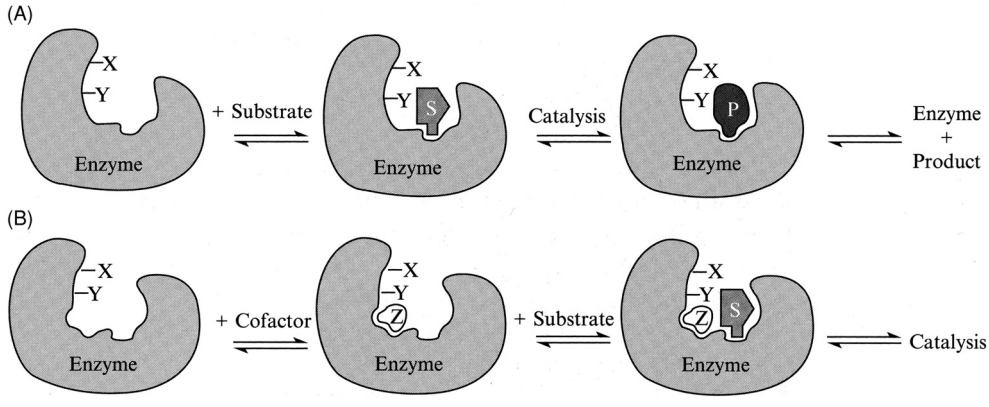


Figure 2.17 *Schematic figure of (A) enzyme + substrate or (B) enzyme + substrate + cofactor.*

enantiomer of the substrate over the other, just as a hand fits a glove. There are four types of enzyme-substrate interactions used by enzymes, as follows:

1) *Electrostatic Interactions*. Substrates containing ionisable functional groups which are charged in aqueous solution at or near pH 7 are often bound via electrostatic interactions to oppositely charged amino acid sidechains at the enzyme active site. Thus, for example, carboxylic acids (pK_a 4-5) are found as the negatively charged carboxylate anion at pH 7, and are often bound to positively charged sidechains such as the protonated ε-amino sidechain of a lysine or the protonated guanidine sidechain of arginine, shown in Figure 2.18.

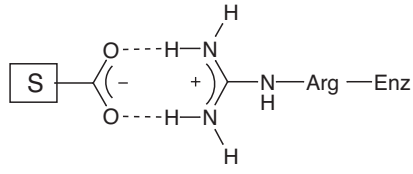


Figure 2.18 *Binding of substrate carboxylate group by an arginine sidechain, involving electrostatic and hydrogen-bonding interactions.*

Similarly, positively charged substrate groups can be bound electrostatically to negatively charged amino acid sidechains of aspartate and glutamate. Energetically speaking, the binding energy of a typical electrostatic interaction is in the range 25–50 kJ mol^{-1}, the strength of the electrostatic interaction varying with $1/r^2$, where r is the distance between the two charges.

2) *Hydrogen Bonding*. Hydrogen bonds can be formed between a hydrogen bond donor containing a lone pair of electrons and a hydrogen bond acceptor containing an acidic hydrogen. These interactions are widely used for binding polar substrate functional groups. The strength of hydrogen bonds depends upon the chemical nature and the geometrical alignment of the interacting groups. Studies of enzymes in which hydrogen-bonding