Introduction to Enzyme and Coenzyme Chemistry

Introduction to Enzyme and Coenzyme Chemistry

Second Edition

TIM BUGG
Professor of Biological Chemistry, Department of Chemistry,
University of Warwick, UK

Editorial offices:
Blackwell Publishing Ltd, 9600 Garsington Road, Oxford OX4 2DQ, UK
Tel: +44 (0)1865 776868
Blackwell Publishing Inc., 350 Main Street, Malden, MA 02148-5020, USA
Tel: +1 781 388 8250
Blackwell Publishing Asia Pty Ltd, 550 Swanston Street, Carlton, Victoria 3053, Australia
Tel: +61 (0)3 8359 1011

First published 1997 by Blackwell Science
Second edition published 2004 by Blackwell Publishing

Library of Congress Cataloging-in-Publication Data

Bugg, Tim.
Introduction to enzyme and coenzyme chemistry / Tim Bugg.–2nd ed.
p. cm.
Includes bibliographical references and index.
ISBN 1-4051-1452-5 (pbk. : alk. paper)
1. Enzymes. 2. Coenzymes. I. Title.

QP601.B955 2004
572′.7–dc22
2003025117

ISBN 1-4051-1452-5

A catalogue record for this title is available from the British Library

Set in 10/13 pt Times
by Kolam Information Services Pvt. Ltd, Pondicherry, India
Printed and bound in Great Britain
by Ashford Colour Press, Gosport

The publisher's policy is to use permanent paper from mills that operate a sustainable forestry policy, and which has been manufactured from pulp processed using acid-free and elementary chlorine-free practices. Furthermore, the publisher ensures that the text paper and cover board used have met acceptable environmental accreditation standards.

For further information on Blackwell Publishing, visit our website:
www.blackwellpublishing.com

Contents

Preface

Since the publication of the first edition in 1998, the field of chemical biology has, I would say, become more a part of the core research and teaching of Chemistry Departments around the world. As the genes and proteins involved in important biological problems are elucidated, so they become accessible for study at the molecular level by chemists. Enzymology is a core discipline within chemical biology, since enzymes are the biological catalysts that make things happen within cells: they translate gene sequence into biological function.

The main feature of the first edition that I wanted to improve was the figures. My original intention was to have a case study in each chapter, where I would illustrate the chemical mechanism, with curly arrows, and on the facing page would show the three-dimensional structure of the enzyme active site, and/or tertiary structure of the enzyme. One of the fascinations of enzymology is the interplay between structure and function, and I wanted to try to convey this to students. In the first edition I was only able to include a set of colour plates in the middle of the book. In the second edition, I have prepared a series of two-colour pictures using Rasmol, to accompany the text. I hope that these pictures convey the desired ideas.

I have also updated the first edition with recent observations and references from the literature, and have added a few new topics. I have written a new chapter entitled Radicals in Enzyme Catalysis (Chapter 11), which includes the discovery of protein radicals and the discovery of SAM-dependent radical reactions. I have also mentioned the recent proposals for protein dynamics (Chapter 3) and proton tunnelling (Chapter 4) in enzyme catalysis, in a way that I hope will be accessible to undergraduate students.

I would like to thank my colleagues, researchers, and students at Warwick for their encouragement, and hope that the book is useful to chemical biologists everywhere.

Tim Bugg
University of Warwick

Representation of Protein Three-Dimensional Structures

In the second edition I have used the program Rasmol to draw representations of protein three-dimensional structures. Rasmol was developed by Roger Sayle (GlaxoSmithKline Pharmaceuticals), is freely available from the internet (http://www.umass.edu/microbio/rasmol), and can be downloaded with instructions for its use. There are several packages available for representation of protein structures, but Rasmol is straightforward to learn and freely available.

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 try downloading a few of these, and viewing each one on a computer screen, as you can turn the structure around and hence 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.

Once you have downloaded the PDB file, then you run the Rasmol program, and open the PDB structure file to view the structure. You can view the protein backbone in several different ways: as individual atoms (wireframe), protein backbone (backbone), strands or cartoons. 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. 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 you can prepare your own multi-colour pictures!

Further reading

R.A. Sayle & E.J. Milner-White (1995) RasMol: Biomolecular graphics for all. *Trends Biochem. Sci.*, **20**, 374–6.

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 AD, 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 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

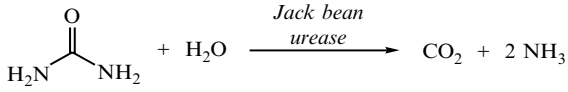


Figure 1.1 Reaction catalysed by the enzyme urease.

1926 of the enzyme urease from Jack beans by Sumner proved beyond doubt that biological catalysis was carried out by a chemical substance (Figure 1.1).

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 acid (DNA)).

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 that 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 used 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 1768–76. 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 B_{12}, 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 coenzymes: small organic cofactors 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 coenzymes that we are going to encounter in this book.

1.4 The commercial importance of enzymes in biosynthesis and biotechnology

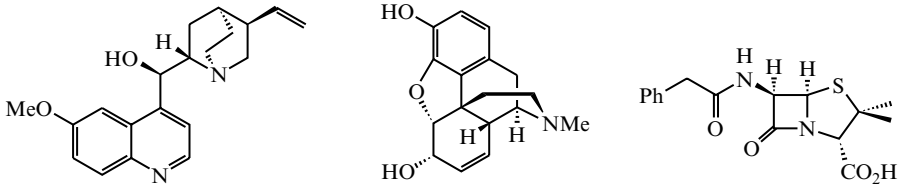
Many plants and micro-organisms contain natural products that possess potent biological activities. The isolation of these natural products has led to the

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 anaemia	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

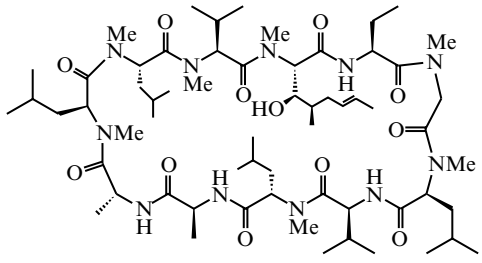
CoA, coenzyme A; FAD, flavin adenine dinucleotide; FMN, flavin mononucleotide; NAD, nicotinamide adenine dinucleotide; PLP, pyridoxal-5′-phosphate; TPP, thiamine pyrophosphate.

discovery of many biologically active compounds such as quinine, morphine and penicillin (Figure 1.2) which have been fundamental to the development of modern medicine.

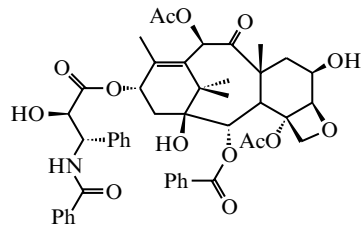


quinine morphine penicillin G

Figure 1.2 Structures of quinine, morphine and penicillin G.



cyclosporin A



taxol

Figure 1.3 Structures of cyclosporin A and taxol.

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 anti-cancer drug isolated from yew bark (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 biosynthesise these molecules with apparent ease using enzyme-catalysed biosynthetic pathways. Hence, there is considerable interest in elucidating the biosynthetic pathways for important natural products 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 (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

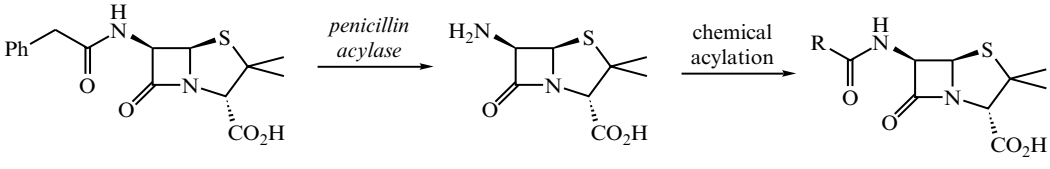


Figure 1.4 Industrial production of a semi-synthetic penicillin using penicillin acylase.

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 organism 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

Table 1.2 Commercial applications of enzyme inhibitors.

Anti-bacterial 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.
Anti-fungal 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.
Anti-viral 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 Section 8.5).

AZT, 3′-azido,3′-deoxythymidine; EPSP, 5-enolpyruvyl-shikimate-3-phosphate.

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.

Further reading

Historical development of enzymology

T.D.H. Bugg (2001) The development of mechanistic enzymology in the 20th century. *Nat. Prod. Reports*, **18**, 465–93.

Enzymes in biosynthesis and biotechnology

J. Mann (1987) *Secondary Metabolism*, 2nd edition. Clarendon Press, Oxford.
C.H. Wong & G.M. Whitesides (1994) *Enzymes in Synthetic Organic Chemistry*. Pergamon, Oxford.

Medicinal chemistry

G.L. Patrick (2001) *An Introduction to Medicinal Chemistry*, 2nd edition. OUP, Oxford.
R.B. Silverman (2001) *The Organic Chemistry of Drug Design and Drug Action*. Academic Press, San Diego.

2 All Enzymes are Proteins

2.1 Introduction

Enzymes are giant molecules. Their molecular weight varies from 5000 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.

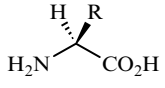
Enzymes belong to a larger biochemical family of macromolecules known as proteins. The common feature of proteins is that they are polypeptides: their structure is made up of a linear sequence of α-amino acid building blocks joined together by amide linkages. This linear polypeptide chain then 'folds'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

2.2 The structures of the L-α-amino acids

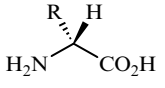
Proteins are composed of a family of 20 α-amino acid structural units whose general structure is shown in Figure 2.1. α-Amino acids are chiral molecules: that is, their mirror image is not superimposable upon the original molecule.

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. 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 20, whose structures are shown in Figure 2.2. The differences between these 20 lie in the nature of the side chain R. The simplest amino acids are glycine (abbreviated Gly or simply G), which has no side chain, and alanine (Ala or A), whose side chain is a methyl group. A number of side chains are hydrophobic ('water-hating') in character, for example the thioether of methionine (Met); the branched aliphatic side chains of valine (Val), leucine (Leu) and isoleucine (Ile); and the aromatic



general structure of an L-α-amino acid



general structure of a D-α-amino acid

Figure 2.1 General structure of L- and D-amino acids.

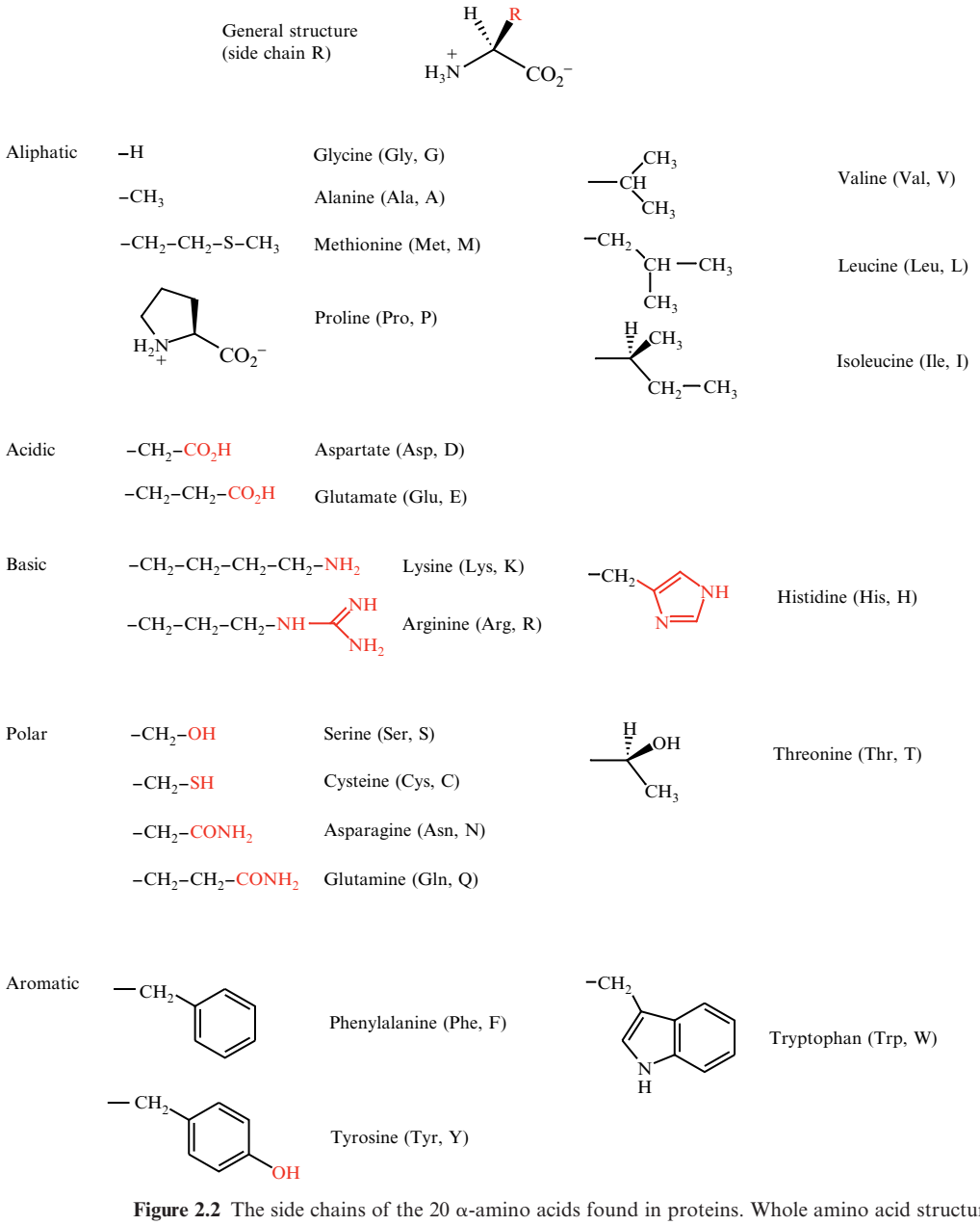


Figure 2.2 The side chains of the 20 α-amino acids found in proteins. Whole amino acid structure shown for proline. Functionally important groups highlighted in red.

side chains of phenylalanine (Phe) and tryptophan (Trp). The remainder of the amino acid side chains are hydrophilic ('water-loving') in character. Aspartic acid (Asp) and glutamic acid (Glu) contain carboxylic acid side chains, and their corresponding primary amides are found as asparagine (Asn) and glutamine

(Gln). There are three basic side chains 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 side chains 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 side chain confers certain physical and chemical properties upon the corresponding amino acid, and upon the polypeptide chain in which it is located. The amino acid side chains 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 100–1000 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.

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

Genes are composed of four deoxyribonucleotides (or 'bases'): deoxyadenine (dA), deoxycytidine (dC), deoxyguanine (dG) and deoxythymidine (dT),

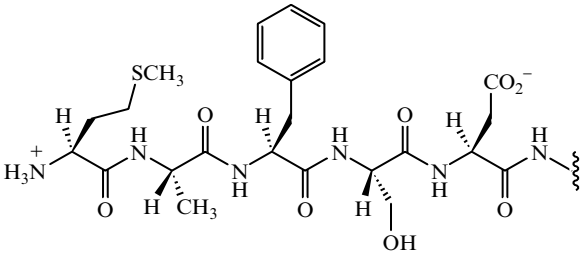


Figure 2.3 Structure of the N-terminal portion of a polypeptide chain.

arranged in a specific linear sequence. To give some idea of size, a typical gene might consist of a sequence of 1000 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 deoxyribonucleotide sequence of the DNA strand is transcribed into messenger ribonucleic acid (mRNA) containing the corresponding ribonucleotides 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.

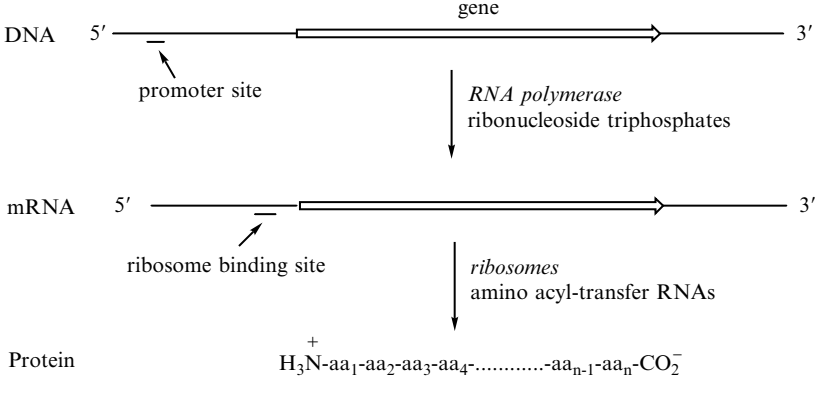


Figure 2.4 Pathway for protein biosynthesis.

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.

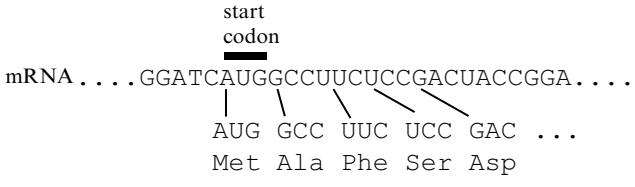


Figure 2.6 Translation of mRNA into protein.

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 onto the next generation.

So if we look at an alignment of amino acid sequences of 'related' enzymes from different organisms, we would expect that catalytically important amino

```
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.

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 side chains 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 this 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 100–1000 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:

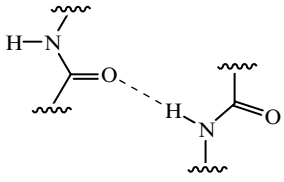


Figure 2.8 A hydrogen bond.

(1) packing of hydrophobic amino acid side chains 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.

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.

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 side chains point outwards from the pitch of the helix, consequently amino acid side chains that are four residues apart in the primary sequence will end up close in space. Interactions between such side chains 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 side chains of the amino acid residues. In Figure 2.10b, the same helix is drawn in 'ribbon' form, a convenient representation that is used for drawing protein structures.

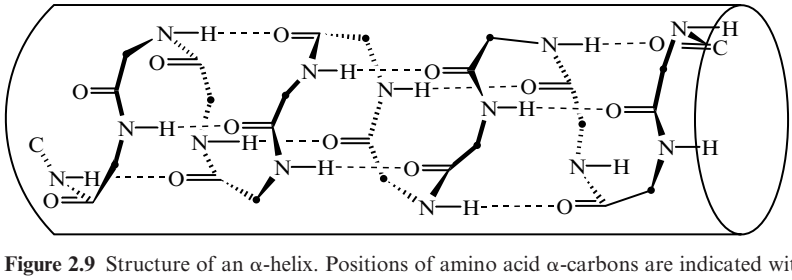


Figure 2.9 Structure of an α-helix. Positions of amino acid α-carbons are indicated with dots.

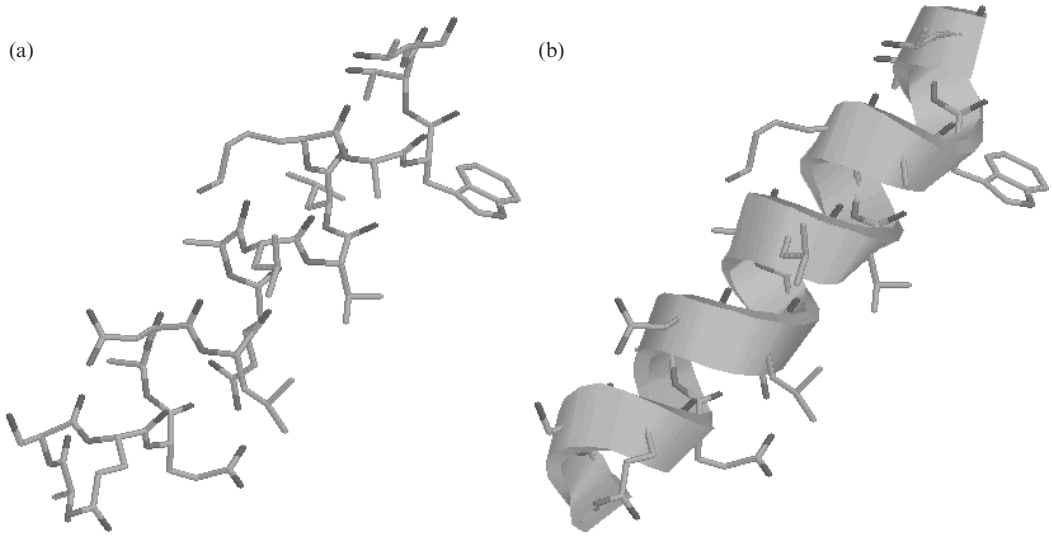


Figure 2.10 Structure of an α-helix, (a) showing positions of the polypeptide chain and side chains and (b) showing the same structure in ribbon format.

The β-sheet is a structure formed by two or more linear polypeptide strands, held together by a series of interstrand 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 anti-parallel, 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 anti-parallel β-sheets in a protein structure, with Figure 2.12b showing the same β-sheets in 'ribbon' form.

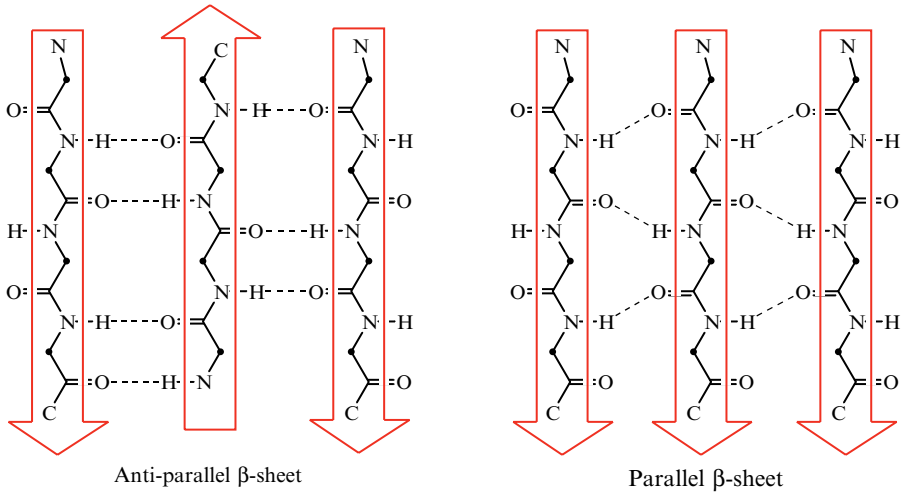


Figure 2.11 Structure of β-sheets. Positions of amino acid α-carbons are indicated with dots.

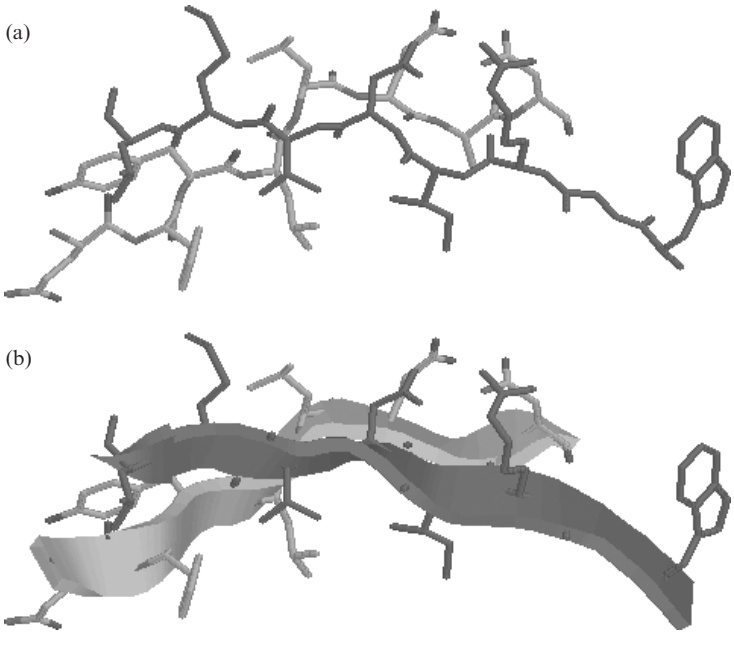


Figure 2.12 Structure of two anti-parallel β-sheets, (a) showing positions of the polypeptide chain and side chains and (b) showing the same structure in ribbon format.

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 sub-units, 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

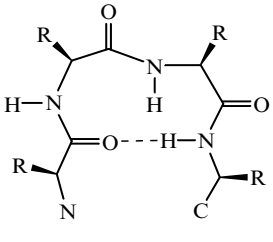


Figure 2.13 Structure of a β-turn.

hydrophobic amino acid side chains on the inside of the protein and positioning hydrophilic amino acid side chains 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) anti-parallel 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 haem-containing cytochromes which act as electron carriers (see Chapter 6) are four-helix 'bundles', illustrated in Figure 2.14 in the case of cytochrome b_{562}. The family of globin oxygen carriers, including haemoglobin, 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. Anti-parallel β structures consist of regular arrays of β-sheet–β-turn–*anti-parallel* β-sheet. For example, the metallo-enzyme superoxide dismutase contains a small barrel of anti-parallel 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 sub-units can bind together to form a homodimer (two identical sub-units), trimer or tetramer, or in other cases non-identical sub-units fit together to form highly complex quaternary structures. One familiar example is the mammalian oxygen transport protein haemoglobin, which consists of a tetramer of identical 16-kDa sub-units.

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 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.

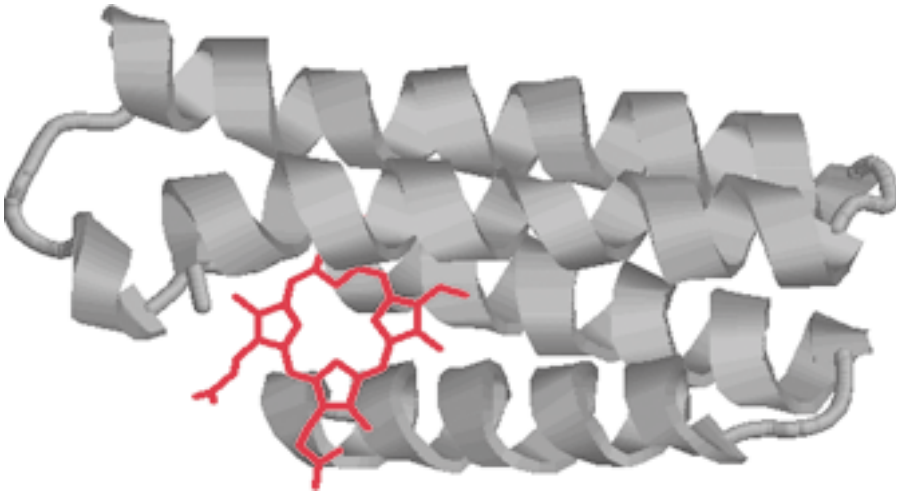


Figure 2.14 Structure of cytochrome b_{562} (PDB file 256B), a four-helix bundle protein. Haem cofactor shown in red.

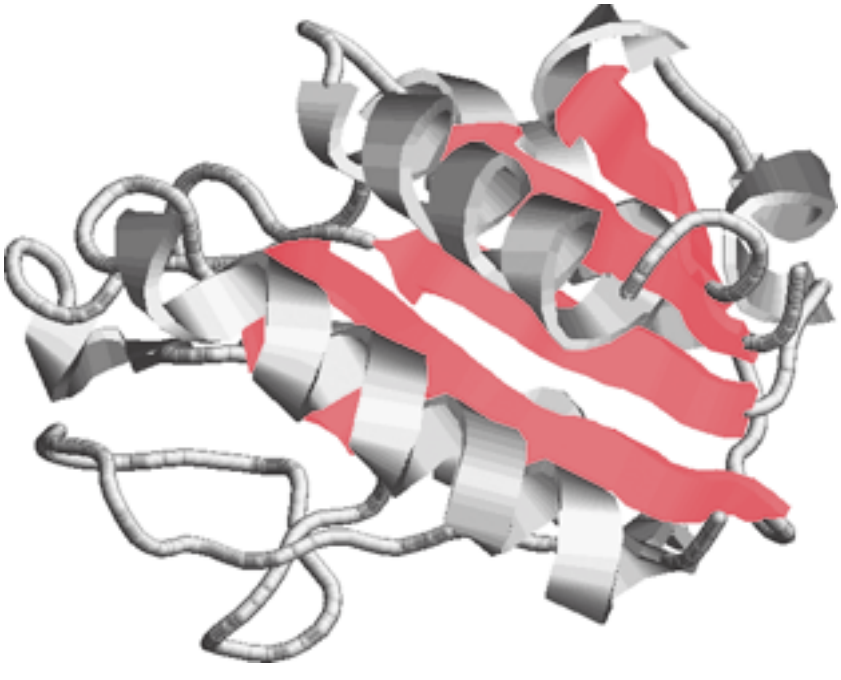


Figure 2.15 Structure of flavodoxin (PDB file 1AHN), a redox carrier protein containing five parallel β-sheets, each connected by an intervening α-helix. Parallel β-sheets shown in red.

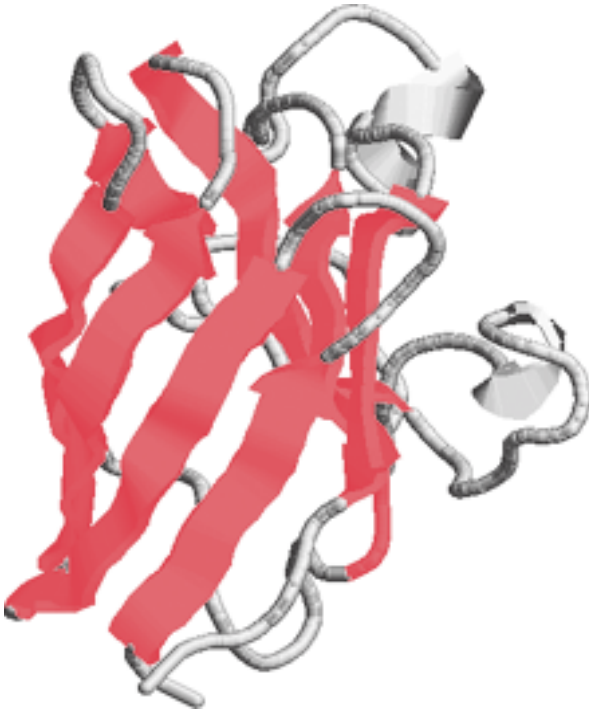


Figure 2.16 Structure of superoxide dismutase (PDB file 1CB4), a β-barrel protein containing eight anti-parallel β-sheets. Anti-parallel β-sheets shown in red.

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 side chains which bind the substrate and carry out the enzymatic reaction, as shown in Figure 2.17a. 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 2.17b.

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 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:

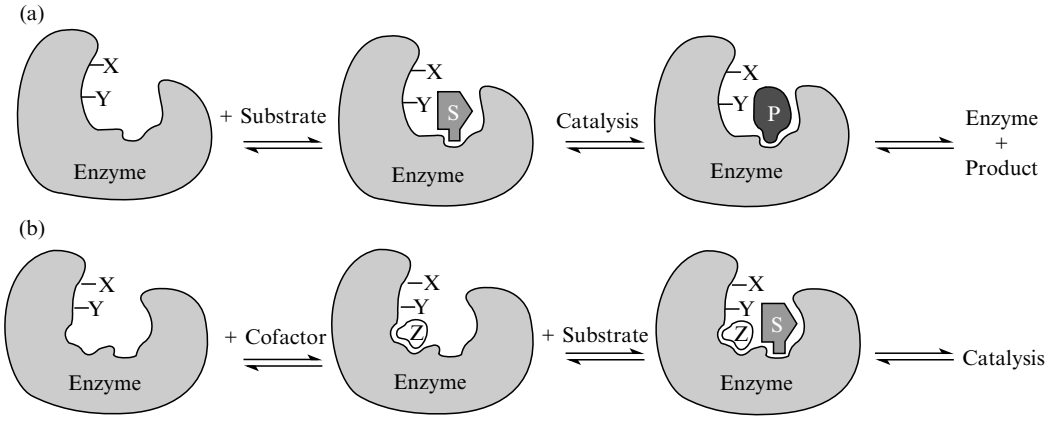


Figure 2.17 Schematic figure of (a) enzyme plus substrate and (b) enzyme plus substrate plus cofactor.

(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 side chains 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 side chains such as the protonated ε-amino side chain of a lysine or the protonated guanidine side chain of arginine, shown in Figure 2.18.

Similarly, positively charged substrate groups can be bound electrostatically to negatively charged amino acid side chains 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 groups have been specifically mutated has revealed that hydrogen

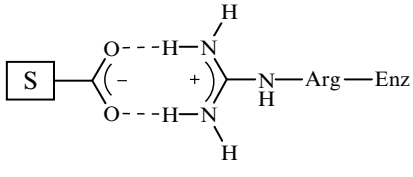


Figure 2.18 Electrostatic enzyme–substrate interaction.