1
Enzymes

RICHARD O. JENKINS
De Montfort University, UK

1.1 Introduction

Enzymes are biological catalysts that mediate virtually all of the biochemical reactions that constitute metabolism in living systems. They accelerate the rate of chemical reaction without themselves undergoing any permanent chemical change, i.e. they are true catalysts. The term 'enzyme' was first used by Kühne in 1878, even though Berzelius had published a theory of chemical catalysis some 40 years before this date, and comes from the Greek *enzumé* meaning 'in (*en*) yeast (*zumé*)'. In 1897, Eduard Büchner reported extraction of functional enzymes from cells. He showed that a cell-free yeast extract could produce ethanol from glucose, a biochemical pathway now known to involve 11 enzyme-catalysed steps. It was not until 1926, however, that the first enzyme (urease from Jackbean) was purified and crystallised by James Sumner of Cornell University, who was awarded the 1947 Nobel Prize. The prize was shared with John Northrop and Wendell Stanley of the Rockefeller Institute for Medical Research, who had devised a complex precipitation procedure for isolating pepsin. The procedure of Northrop and Stanley has been used to crystallise several enzymes. Subsequent work on purified enzymes, by many researchers, has provided an understanding of the structure and properties of enzymes.

All known enzymes are proteins. They therefore consist of one or more polypeptide chains and display properties that are typical of proteins. As considered later in this chapter, the influence of many chemical and physical parameters (such as salt concentration, temperature and pH) on the rate of enzyme catalysis can be explained by their influence on protein structure. Some enzymes require small non-protein molecules, known as cofactors, in order to function as catalysts.

Enzymes differ from chemical catalysts in several important ways:

1. Enzyme-catalysed reactions are at least several orders of magnitude faster than chemically-catalysed reactions. When compared to the

corresponding uncatalysed reactions, enzymes typically enhance the rates by 10^6 to 10^{13} times.
2. Enzymes have far greater reaction specificity than chemically-catalysed reactions and they rarely form byproducts.
3. Enzymes catalyse reactions under comparatively mild reaction conditions, such as temperatures below 100°C, atmospheric pressure and pH around neutral. Conversely, high temperatures and pressures and extremes of pH are often necessary in chemical catalysis.

1.1.1 In this chapter

This chapter is concerned mainy with the fundamental aspects of enzymes that determine their properties and catalytic capabilities. It is intended to provide a sound basis for understanding of many of the applied aspects of enzymes considered in subsequent chapters in this text. Given the wealth of fundamental knowledge on enzymes, it is only possible here to provide a perspective on each of the topics. Some of the topics will be considered in more detail, or from a different perspective, later on in the text.

Section 1.2 deals with the classification and nomenclature of enzymes. It considers some of the rules that form the basis of a rational system classification and naming enzymes, and provides examples of enzymes in each of the six main classes. Much of the chapter is devoted to protein structure (Section 1.3) because this ultimately defines the properties of enzymes, such as substrate specificity, stability, catalysis and response to physical and chemical factors. Protein structure is considered at all levels of organisation, from the 'building blocks' (amino acids) of proteins, through backbone conformations and three-dimensional shapes, to enzymes having more than one sub-unit. Consideration of the forces that stabilise protein molecules follows (Section 1.4) and the strengths of the various bonds are compared in relation to level of protein structure. Section 1.5 briefly describes some of the basic properties of proteins, such as chemical reactions with reactive amino acid groups, the acid–base properties of enzymes and some other factors (temperature and pH) that influence protein solubility and catalytic activity. Cellular biosynthesis of proteins is described in Section 1.6, with the emphasis very much on the process of reading the genetic code to synthesising a chain of amino acids in the correct predetermined sequence. This is followed by a section (1.7) on enzymatic modification of proteins within cells *after* they have been synthesised. Such post-translational modification influences the structural stability or activity of enzymes. Section 1.8 considers enzymatic catalysis, with the emphasis on enzyme substrate specificity and the requirement of some enzymes for the presence of non-proteinaceous compounds for catalytic activity. Comments on future trends (Section 1.9) and recommendations for further reading (Section 1.10) are

also included. Papers from the primary literature have not been referred to; rather, a list of relevant books and review articles are provided at the end of the chapter.

1.2 Classification and nomenclature of enzymes

Organisms – whether animal, plant or microorganism – are both complex and diverse. In biological systems, thousands of different types of reactions are known to be catalysed by different enzymes; many more are yet to be discovered. The diversity of enzymes is, therefore, enormous in terms of type of reaction(s) they catalyse, and also in terms of structure. Enzymes range from individual proteins with a relative molecular mass (RMM) of around 13 000 catalysing a single reaction, to multi-enzyme complexes of RMM several million catalysing several distinct reactions.

Given such diversity, it is essential to have a rational basis for classification and naming of enzymes. Currently, it is the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB) that considers these matters and gives recommendations to the international scientific community.

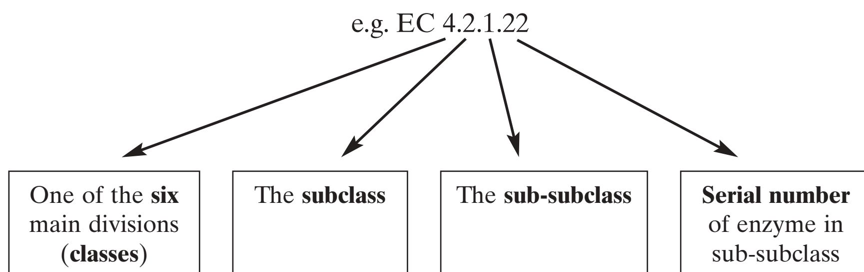
1.2.1 General rules

Enzymes are principally classified and named according to the **chemical reaction** they catalyse, as this is the specific property that distinguishes one enzyme from another. It is the observed chemical change produced by the complete enzyme reaction that is used for this purpose, i.e. the overall reaction, rather than the formation of intermediate complexes of the reactants with the enzyme. Some notable consequences of this system are:

- A systematic name cannot be given to an enzyme until the chemical reaction is known. This applies, for example, to enzymes that catalyse an isotopic change to a molecule that indicates one step in the overall reaction, but the reaction as a whole remains unknown.
- An enzyme name is assigned not to a single enzyme protein but to a group of proteins with the same catalytic property. Some exceptions exist, where more than one name is assigned to enzymes with the same catalytic property because the reaction is so different in terms of substrate specificity or mechanism. Other exceptions include acid and alkaline phosphatases. These enzymes carry out the same reaction but at widely different pH values.
- Enzymes from different sources – such as animal, plant and microorganisms – are classified as one entry.

- To classify an enzyme it is occasionally necessary to choose between alternative ways of regarding the chemical reaction. In general, the alternative selected should reduce the number of exceptions.
- The direction of the chemical reaction needs to be considered, since all reactions catalysed by enzymes are reversible. For simplicity, the direction chosen should be the same for all enzymes in a given class even if this direction has not been shown for all of the enzymes.

The Enzyme Commission of the International Union of Biochemistry, in its report of 1961, devised a rational system for classification of enzymes and assigning code numbers to them based on the reaction catalysed. The code numbers, prefixed by EC, are now used widely and contain four elements separated by points:



There are six classes of enzymes that are distinguished by the first digit of the EC code (Table 1.1). The second and third digits describe further the type of reaction catalysed. These digits are defined for each of the separate main classes of enzymes and there is no general rule that applies to their meaning. Enzymes that catalyse very similar reactions, e.g. enzymes that cleave C—O bonds in a substrate molecule, will have the same first three digits in their EC code. They will, however, have different fourth digits that define the actual substrate for the reaction.

A consequence of enzymes being classified according to the chemical reaction they catalyse is that **isoenzymes** (different enzymes catalysing identical reactions) carry the same four digit EC classification number. There are, for example, five different isoenzymes of lactate dehydrogenase in the human body and the EC code does not provide a means of distinguishing between them. Rather, the particular isoenzyme and its source (e.g. mammalian heart) have also to be specified.

1.2.2 Recommended and systematic names

The Enzyme Commission has recommended that there should be 'systematic' as well as 'trivial' (working) nomenclatures for enzymes; examples for

Table 1.1 Classification and nomenclature for the six classes of enzymes

1. **Oxidoreductases**: enzymes that catalyse oxidoreductase reactions
   2nd EC digit: indicates group in the hydrogen donor (substrate oxidised), e.g. —CHOH—, aldehyde, keto
   3rd EC digit: indicates type of acceptor involved, e.g. a cytochrome, molecular oxygen, an iron–sulphur protein, etc
   Systematic name: *donor:acceptor oxidoreductase*
   Recommended name: *donor:dehydrogenase* (*reductase* as alternative; *oxidase* where O_2 is acceptor)
   e.g. alcohol dehydrogenase (trivial); alcohol NAD^+ oxidoreductase (EC 1.1.1.1)
2. **Transferases**: enzymes transferring a group
   2nd EC digit: indicates group transferred, e.g. methyl, glycosyl, phosphate
   3rd EC digit: further information on group transferred, e.g. hydroxymethyl
   Systematic name: *donor:acceptor grouptransferase*
   Recommended name: *acceptor grouptransferase*
   e.g. glucokinase; ATP glucose phosphotransferase (EC 2.7.1.2)
3. **Hydrolases**: enzymes that catalyse cleavage of C—O, C—N, C—C and some other bonds
   2nd EC digit: indicates nature of bond hydrolysed, e.g. ester, glycosyl
   3rd EC digit: indicates nature of substrate, e.g. carboxylic ester, thiolester
   Systematic name: *substrate:hydrolase*
   Recommended name: *substrate* with suffix *-ase*
   e.g. carboxypeptidase A (EC 3.4.17.1)
4. **Lyases**: enzymes that cleave C—C, C—O, C—N and other bonds by elimination, leaving double bonds or rings, or add groups to double bonds
   2nd EC digit: indicates the bond broken
   3rd EC digit: further information on group eliminated, e.g. CO_2, H_2O
   Systematic name: *substrate group-lyase* (hyphen included)
   Recommended names: e.g. *decarboxylase*, *dehydratase* (in case of elimination of CO_2 and H_2O); *synthase* used if reverse reaction described
   e.g. pyruvate decarboxylase; pyruvate-lyase (EC 4.1.1.1)
5. **Isomerases**: enzymes that catalyse geometric or structural changes within one molecule
   2nd EC digit: indicates type of isomerism, e.g. racemase, epimerase, *cis-trans* isomerase
   3rd EC digit: indicates type of substrate
   Systematic name: *substrate:type of isomerism*
   Recommended name: *substrate:isomerase*
   e.g. maleate isomerase; maleate *cis-trans* isomerase (EC 5.2.1.10)
6. **Ligases**: enzymes catalysing the joining of two molecules coupled with hydrolysis of a diphosphate bond in ATP (or similar triphosphate)
   2nd EC digit: indicates the bond formed, e.g. C—O, C—S, C—N
   3rd EC digit: (only used in the C—N ligases)
   Systematic name: *X:Y ligase (ADP-forming)*
   Recommended name: *X:Y ligase* (previously *synthetase* was used)
   e.g. pyruvate carboxylase (trivial); pyruvate carboxyligase (ADP forming) (EC 6.4.1.1)

each of the six classes of enzymes are given in Table 1.1. The systematic name describes the action of an enzyme as exactly as possible, whereas the trivial name is sufficiently short for general use and is often a name already in common use. The Enzyme Commission-recommended trivial names for new enzymes are often condensed versions of systematic names.

Since enzymes are divided into groups according to the type of reaction catalysed, this and the name(s) of the substrate(s) are the basis for systematic naming of individual enzymes. It is also the basis for classification and code numbers. Names of enzymes, especially those ending in *ase*, generally refer to single enzymes and are not applied to systems containing one or more enzymes. When an overall reaction involving more than one enzyme is named, the word 'system' is included in the name. For example, the 'succinate oxidase system' is used to describe the enzymatic oxidation of succinate involving succinate dehydrogenase, cytochrome oxidase and several intermediate carriers.

General rules for systematic names and guidelines for recommended names, as well as rules and guidelines for particular classes of enzymes, are available at Enzyme Nomenclature Database at the Swiss Institute of Bioinformatics (http://www.espasy.ch/enzyme).

1.3 Protein structure

1.3.1 Overview

Proteins consist of one or more polypeptides and each polypeptide is a chain of amino acids linked together by peptide bonds. A different gene codes for each polypeptide and determines the sequence of amino acids of the polypeptide. Polypeptide chains fold up when synthesised to form a unique three-dimensional shape (conformation), determined by their amino acid sequences. Multiple weak interactions stabilise the conformation of polypeptides and factors (such as pH, heat and chemicals) that disrupt these interactions distort the polypeptide's conformation. Enzymes lose their functional activity when their three-dimensional conformation is distorted in this manner, through enzyme denaturation. This demonstrates a clear dependence of enzyme functioning upon protein structure.

There are two main types of proteins: 'fibrous' and 'globular'. **Fibrous** proteins normally have a structural role in biological systems. They are insoluble in water and are physically durable/strong. The three-dimensional structure of fibrous proteins is relatively simple and usually elongated. Examples of fibrous proteins are:

- α-Keratin: the main protein of hair, nails, wool, horn and feathers
- β-Keratin: the main structural component of silk and spider's web

- Collagen: a major protein of cartilage, tendons, skin and bones
- Elastin: a protein found in ligaments in the walls of arteries.

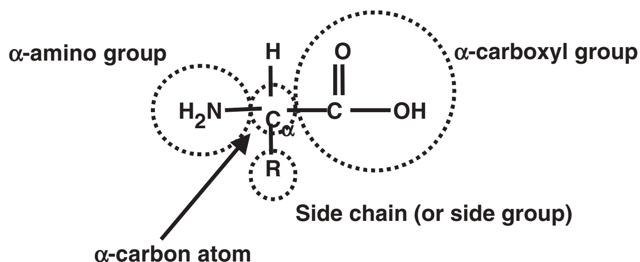
Globular proteins are generally soluble in water and can often be crystallised from solution. They have a more complex three-dimensional structure and tend to adopt an approximate spherical shape in which the amino acid chain is tightly folded. Globular proteins have functional roles in biological systems and all enzymes are globular proteins. Proteins are also categorised as 'simple' or 'conjugated'. **Simple** proteins are composed entirely of amino acids, while **conjugated** proteins contain one or more other materials bound to one or more of the amino acid residues. Examples of conjugated proteins and their bound components are:

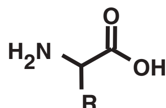
- Glycoproteins – carbohydrate
- Metalloproteins – metal ions
- Lipoproteins – lipids
- Nucleoproteins – nucleic acids
- Flavoproteins – flavin nucleotides

1.3.2 Amino acids – the 'building blocks' of proteins

Amino acids are organic molecules that contain an amino group (primary —NH_2; secondary >NH) and a carboxyl group ($O{=}\overset{|}{C}{-}OH$ or —COOH). There are 20 commonly occurring amino acids. All except one has a central (α) carbon atom, to which is attached a primary amino group (—NH_2), a carboxyl group (—COOH), a hydrogen atom and a side group or chain (R); the side groups are different in all amino acids. Proline is unique because it lacks a primary amino group; instead it contains a secondary amino group (>NH). In proline the side group is curled round so that the nitrogen and the α-carbon atoms form part of a non-polar and fully saturated five-membered imino ring; proline is termed an imino acid. Representations of the generalised structure of amino acids are shown in Fig. 1.1.

Side chains (R-groups) of α-amino acids are polar or non-polar. The structure of the polar molecules may be stabilised by hydrogen bonding in aqueous solution; they display ionic character and are therefore hydrophilic and soluble in water. Conversely, non-polar molecules are relatively insoluble in water, but more soluble in organic solvents. The categorisation of amino acids according to the hydrophobic or hydrophilic character of the side chains is shown in Fig. 1.2. Phenylalanine, tryptophan and tyrosine are termed aromatic amino because the R-group has a six-membered aromatic benzene ring, whereas hystidine has a five-membered imidazole ring. The double-ringed R-group of tryptophan is called indole and in tyrosine the ring is linked to —OH to form a phenolic group. As mentioned earlier,





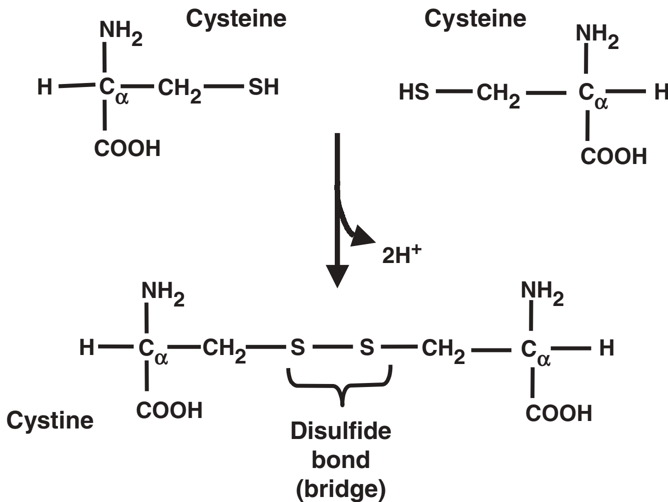
1.1 Representations of the generalised structure of amino acids.

X ⟷ (O, OH, NH_2)

Polar side chains		Non-polar side chains
Uncharged at pH 7	Charged at pH 7	
H—X Glycine (Gly)	Glutamate (Glu) or Glutamic acid	H_3C—X Alanine (Ala)
HO Serine (Ser)	Aspartate (Asp) or Aspartic Acid	H_3C, CH_3 Valine (Val)
HS Cysteine (Cys)	$H_3\overset{+}{N}$ Lysine (Lys)	H_3C, CH_3 Leucine (Leu)
CH_3, HO Threonine (Thr)	NH_2, $H_2\overset{+}{N}$, NH Arginine (Arg)	CH_3, H_3C Isoleucine (Ile)
H_2N, O Asparagine (Asn)		Phenylalanine (Phe)
H_2N, O Glutamine (Gln)		NH Tryptophan (Trp)
H, N, N Histidine (His)		H_3C, S Methionine (Met)
HO Tyrosine (Tyr)		H, N, O, OH Proline (Pro) (complete structure)

1.2 Categorisation of amino acids according to hydrophobic or hydrophilic character of side chains.

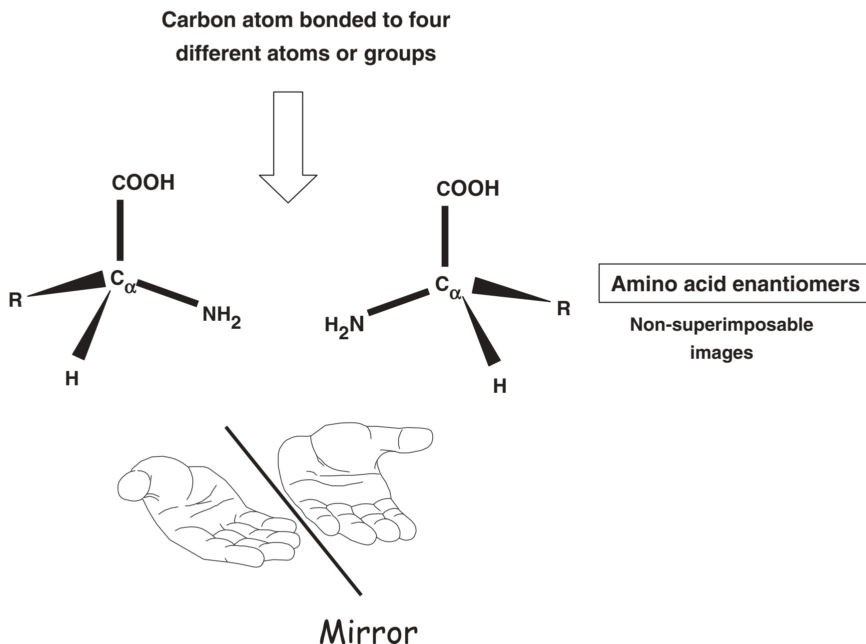
proline – the only other amino acid containing a ring structure – is an imino acid. Several of the hydrophobic side chains are branched-chain aliphatic hydrocarbons. Glutamic acid and aspartic acid have hydrophilic side chains containing carboxyl groups, which are converted to amide groups in



1.3 Oxidation of the sulfydryl groups of cysteine to form cystine.

asparagine and glutamine, respectively. Arginine has a guanidine side chain that, in common with the side chain of lysine, contains an amino group. Cysteine and methionine are sulfur-containing amino acids. In cysteine the sulfydryl group (—SH) oxidises readily to form the dimeric compound cystine, which comprises two cysteine residues linked by a disulfide bridge or bond (Fig. 1.3).

With the exception of glycine, all of the common amino acids exist as optical isomers. These are two mirror image forms of an amino acid that cannot be superimosed by rotation of the molecule. They arise because amino acids (with the exception of glycine) contain at least one α-carbon atom covalently linked to four different atoms or groups. The molecule is therefore asymmetric because no plane drawn through the α-carbon atom can divide the molecule into two parts that are exact mirror images. It follows that two mirror image forms of the complete molecule can exist. The isomers are termed optical isomers because one will rotate the plane of polarized light to the left, and the other to the right. They are termed D- and L-isomers for the sake of distinction, although this does not indicate how they affect the plane of polarised light. Figure 1.4 illustrates the arrangement of groups around an asymmetric α-carbon of an amino acid. Proteins are almost exclusively composed of L-amino acids. The reason for this is that protein biosynthesis is mediated by enzymes that distinguish the optical isomers of amino acids in a solution containing both L- and D-forms (enantiomers); steriospecificity of enzyme action is considered in more detail in Section 1.8.1.

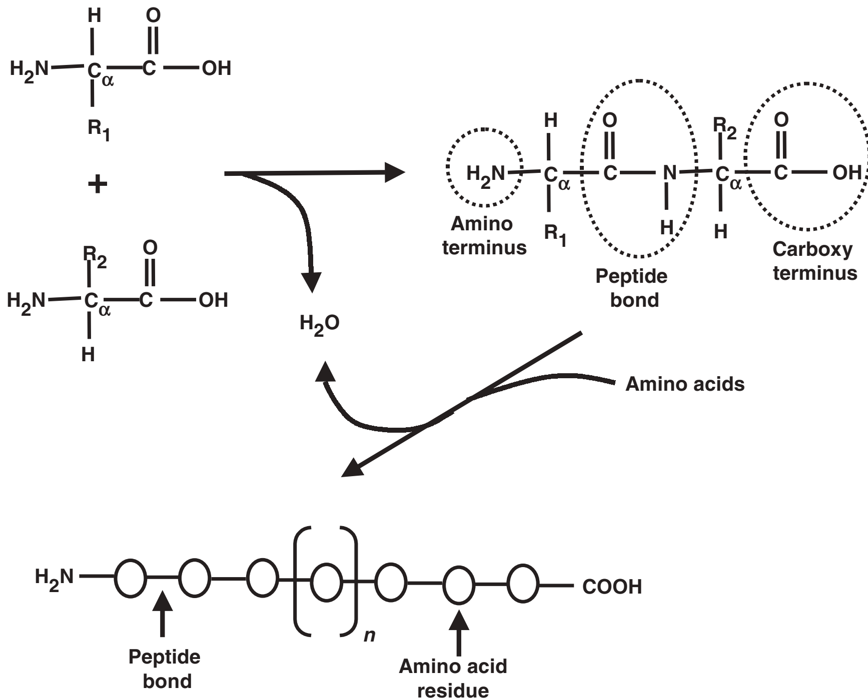


1.4 Arrangement of groups around an asymmetric α-carbon of an amino acid.

1.3.3 Primary structure of proteins

Amino acids (monomers) are joined together by peptide bonds to give proteins. Addition of increasing numbers of amino acids gives peptides and then polypeptides. If the RMM of the chain is more than 5000, the molecule is usually referred to as a polypeptide rather than a peptide. The primary structure of a polypeptide refers to the amino acid sequence, together with the positioning of any disulfide bonds that may be present.

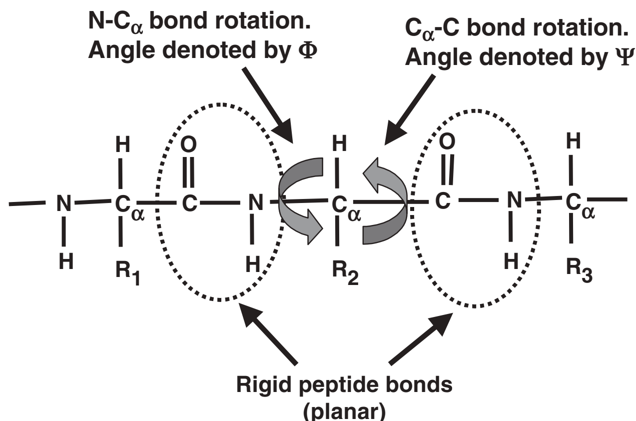
The peptide (amide) bond that joins amino acids together to form a polypeptide is formed by elimination of water, i.e. a condensation reaction (Fig. 1.5). The polypeptide chain formed (the residue) has one free carboxy and one free amino group at opposite ends of the molecule, known as the carboxy and amino termini, respectively. Peptide bonds are rigid, being stabilised by resonance, i.e. the amide nitrogen lone pair of electrons is delocalised across the peptide linkage. The bond can be thought of as having an intermediate form between the two extremes (*cis* and *trans* forms). However, in most instances steric interference between the amino acid side groups and the α-carbon atoms of adjacent amino acid residues means that the *trans* form (R-group lies on opposite sites of the polypeptide chain; Fig. 1.5) is around 1000-fold more common than the *cis* form. This minimizes



1.5 Linking of amino acids by peptide bond formation.

steric interference between the R groups in the peptide chain. In the case of proline, however, the unusual side group of this amino acid allows the peptide bond to adopt the *cis* configuration.

While the peptide bond has a rigid and planar structure, the other bonds in the polypeptide are free to rotate (Fig. 1.6). Rotation about the N—C_{α} bond is denoted by Φ (phi) and the C_{α}—C bond by ψ (psi). When the amino acids are in the *trans* form the polypeptide chain is fully extended and, by convention, these rotation angles (also known as dihedral or torsion angles) are defined as being 180°. In principal, each bond can rotate in either direction and have values –180° and +180°. However, steric hindrance between the atoms of the polypeptide backbone and those of amino acid side chains restricts the degree of rotation and the majority of ψ and Φ combinations are excluded. Rotational freedom around glycine residues is relatively high since steric hindrance is minimised by the small R-group (i.e. hydrogen). Conversely, rotation around proline residues is restricted owing to the unusual structure of the side chain of this amino acid. The ψ and Φ angles, together with the fixed ω angles of all the residues, define the conformation of the main chain (backbone) of the polypeptide. Certain combinations of ψ and Φ form relatively stable regularly-shaped backbone



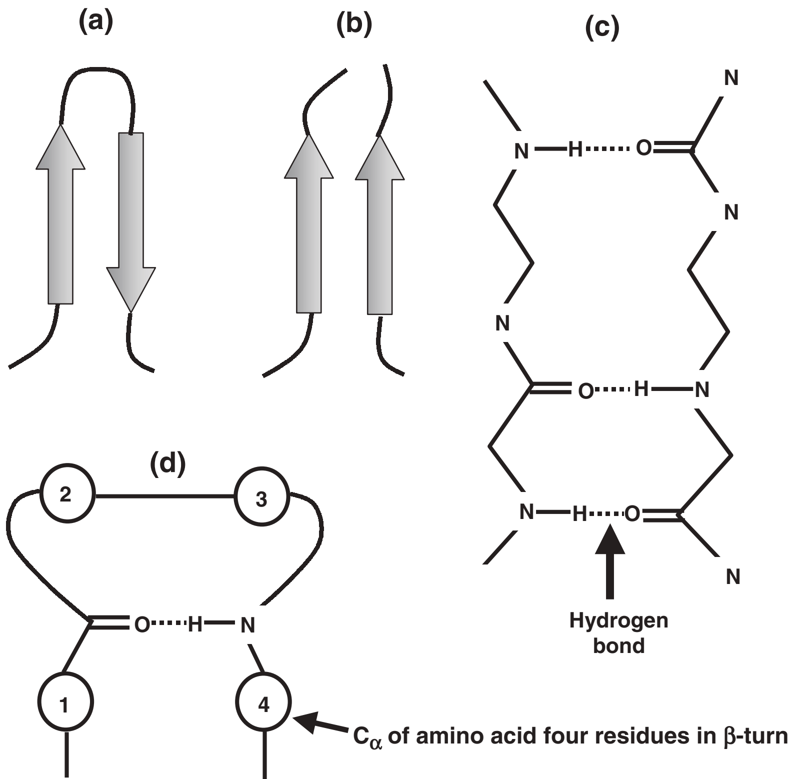
1.6 Rotation in a section of polypeptide chain.

conformations called **secondary structures**. These angles also have a major influence on the final three-dimensional shape (**tertiary structure**) of the polypeptide.

1.3.4 Secondary structure of proteins

Secondary structure is the local spatial conformation of the back bone of a polypeptide, excluding the side chains (R-groups) of the amino acids. Three secondary structures are common in proteins: α-helices, β-sheets and β-turns. These structures are commonly formed because they minimise steric interference between adjacent side-chain groups and maximise formation of intermolecular hydrogen bonds that are closely situated in the primary structure. These bonds stabilise the secondary structures. Parts of the polypeptide backbone that do not have recognisable secondary structures are referred to as 'random coils', or sometimes merely as 'coils', and tend to have more flexibility of movement in solution compared to the secondary structures.

α-Helices are right handed helices containing 3.6 amino acid residues in a full turn. This arrangement is stabilised by hydrogen bonding between carboxyl and amino groups in the polypeptide backbone; every —C═O group forms a hydrogen bond with the —N—H group of the amino acid residue four positions ahead of it in the helix. This means that the hydrogen bonds link together a 13-atom length of polypeptide backbone and thus the α-helice is described as 3.6_{13}. α-Helices are found in both fibrous and globular proteins. In enzymes (globular proteins), the average length of a helical region is three turns, but may vary from a single turn to more than ten consecutive turns. Sometimes a single turn of 3_{10} is found at the end of α-helices.



1.7 β-Sheets made up of two β-strands of the polypeptide chain. Orientation of strands in (a) an anti-parallel and (b) a parallel β-sheet. (c) Hydrogen bonding stabilising an anti-parallel β-sheet. (d) Hydrogen bonding in a β-turn over four amino acid residues.

The side groups of the amino acid residues project outwards, thus the helix has a hollow core.

β-Sheets are relatively extended sections of the polypeptide backbone. They are made up of two β-strands of the polypeptide chain. Each strand is usually five to ten amino acid residues in length, with adjacent peptide groups tilted in alternate directions, giving a zig-zag or pleated conformation; the two strand sections are also termed **β-pleated sheets**. The >N—H and >C=O groups of the amino acid residues that point out at approximate right angles to the extended polypeptide backbone, form hydrogen bonds between two β-strand regions from the same polypeptide or from different polypeptides held in close proximity. Parallel β-sheets occur when the polypeptide chains run in the same direction (i.e. N to C), whereas anti-parallel β-sheets have chains running in opposite directions (Fig. 1.7a,b). The latter structure is the most stable of the two because hydrogen bonding

is more effective in the anti-parallel conformation (Fig. 1.7c). A mixed β-sheet contains both parallel and anti-parallel strands.

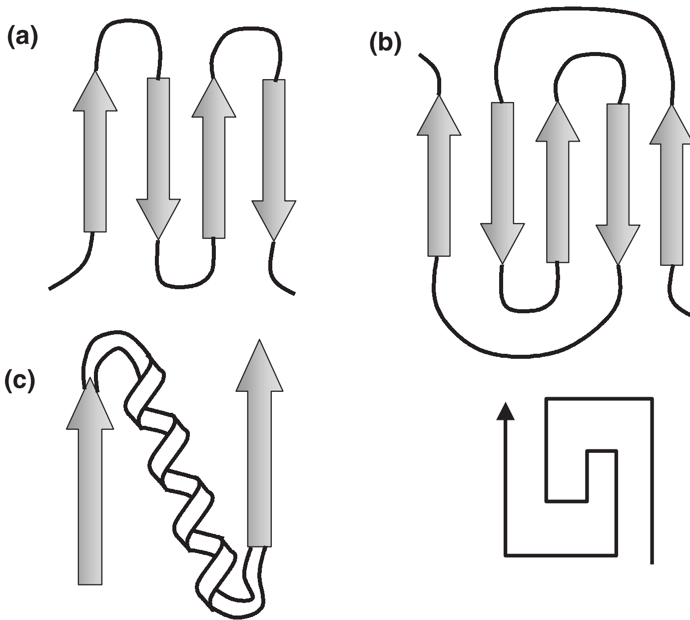
β-Sheets, like α-helices, are found in both fibrous and globular proteins and are essentially linear structures. In order for polypeptides to fold to form a compact tertiary structure, changes in direction of the polypeptide backbone are necessary. Such changes in direction occur in 'loop' regions of the polypeptide between stretches of regular secondary structures (β-sheets and/or α-helices). Loops are generally found on the surface of the polypeptide since they are rich in charged (polar) amino acid residues that bond with surrounding water molecules. In globular proteins, which have a roughly spherical shape, the commonest type of loop structure is the **β-turn** (also known as β-bend or reverse turn). The β-turn introduces a 180° change in direction of the polypeptide chain over four amino acid residues (Fig. 1.7d). Glycine and proline are the prominent amino acids in β-turns and give rise to the change in direction of the polypeptide chain. Proline naturally introduces a twist into the polypeptide chain owing to its ring structure side chain. Glycine, owing to its small side group, minimises steric interference and occupies the restricted space available. A hydrogen bond forms between the —C=O of the first amino acid residue and the NH of the fourth residue, which helps to stabilise the loop (Fig. 1.7d).

1.3.5 Tertiary structure of proteins

The tertiary structure of a protein is the exact three-dimensional shape of the folded polypeptide, i.e. the positioning in space of all the atoms in the polypeptide relative to each other. Polypeptides comprising more than around 200 amino acid residues often have two or more structural sub-units, known as 'domains'. These are tightly folded sub-regions of a single polypeptide, which are connected by more flexible and extended regions of the polypeptide. Domains are usually comprised of 'structural motifs' (sometimes referred to as 'supersecondary structures'), which are secondary structures occurring as closely associated structures. Figure 1.8 illustrate some common structural motifs. The three-dimensional shape of an actual enzyme, glycosyl hydrolase, is illustrated in Fig. 1.9.

Although polypeptides are most often characterised according to their biological activity or function, they can also be characterised based on the domain structure of the polypeptide. There are three main domain types:

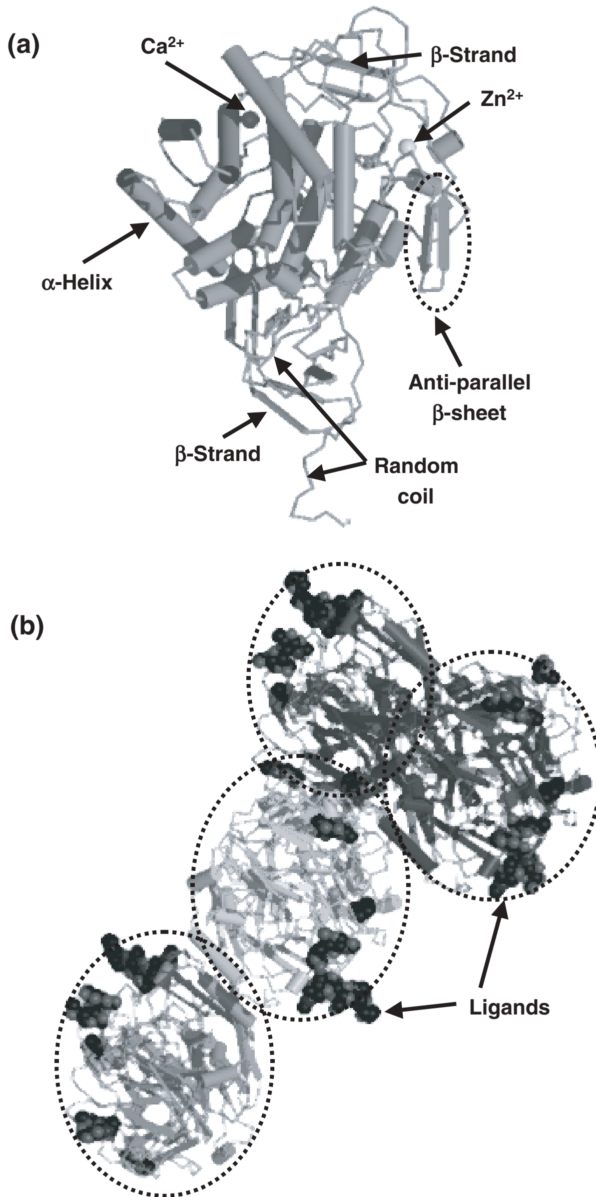
1 α-domain. Comprised exclusively of α-helices, e.g. arranged as a four-helical bundle motif.
2 αβ-domain. Most common domain types in proteins. Comprised of parallel β-sheets surrounded by stretches of α-helix.
3 β-domain. Comprised of a core of anti-parallel β-sheets.



1.8 Some common structural motifs (supersecondary structures) of proteins: (a) β-meander; (b) Greek key (common design in Greek architecture); (c) βαβ unit.

The tertiary structure of a protein is usually so compact that there is virtually no room for water. For non-membrane associated proteins, non-polar side chains of the amino acid residues are nearly all located in the middle of the structure, whereas polar side chains are located on the surface.

X-ray diffraction, nuclear magnetic resonance (NMR) and electron microscopy are techniques used to elucidate the three-dimensional structures of proteins, and each have their advantages and limitations. Electron microscopy provides information only at low-resolution and its principal application has been to provide information on the overall three-dimensional shape of large proteins or protein aggregates. Both X-ray defraction and NMR, however, provide high-resolution information at the atomic level. X-ray diffraction involves bombarding a crystal of a protein with electromagnetic radiation of a wavelength approximately the same as the dimensions of the atoms in a protein, i.e. X-rays, 10^{-10} m. Some of the X-rays are diffracted by the atoms in the protein crystal, giving a diffraction pattern that reflects the three-dimensional structure of the protein molecules in the crystal. This is a powerful technique whose main limitation is the requirement for the protein to be in crystalline form. The vast majority of globular proteins do not crystallise readily; they are generally large, have



1.9 Three-dimensional structure of glycosyl hydrolase (synonym: cellulase endo-1,4-β-glucanase D) from *Clostridium thermocellum*. Main-polypeptide chain: 541 amino acid residues comprising 17α-helices, 51β-strands; associated metals, $3 \times Ca^{2+}$, $1 \times Zn^{2+}$. Data obtained from the Protein Data Bank.[20]

irregular surfaces and tend to hold significant amounts (>30%) of water within their structures.

NMR can be used to determine protein structure in solution. The technique involves subjecting a sample of the protein solution to a strong magnetic field. This causes the spin of the atomic nuclei in the protein (such as ^{13}C and ^{1}H) to align along the magnetic field and, if the appropriate radiofrequency energy is applied, the alignment can be converted to an excited state. When the nuclei then revert back to an unexcited state, radiofrequency radiation is emitted whose exact frequency is influenced by the molecular environment of the individual nuclei. Detection and measurement of the emitted radiation therefore provides information on the three-dimensional structure of proteins. The main limitation of NMR in this application is the complexity of the data that is generated and the technique is generally applied only to relatively small proteins, having a relative molecular mass of <25 000.

1.3.6 Quaternary structure of proteins

Some proteins are made up of more than one polypeptide chain, or subunit (Fig. 1.9). The quaternary structure of a protein describes subunit structure. Proteins that lack quaternary structure are referred to as **monomeric**, those having two-subunits as **dimeric**, three as **trimeric**, four as **tetrameric**, etc. Different types of subunits are assigned Greek letters, with the number of each subunit indicated by a subscripted number. For example, in the case of phosphorylase kinase, the subunit composition is described as $\alpha_4\beta_4\gamma_4\delta_4$, indicating that there are four each of four different subunits, i.e. 16 separate polypeptides.

The vast majority of enzymes are oligomeric enzymes, with molecular weight in excess of 35 000. Some oligomeric enzymes form **isoenzymes**, which are different enzymes catalysing the same chemical reaction; they carry the same four digit EC classification number (see Section 1.2). Isoenzymes typically have different kinetic properties and commonly arise through the occurrence of alternative subunits in an oligomeric enzyme. For example, mammalian lactate dehydrogenase is a tetramer, in which the subunits may be of two types: the heart-type (H) and the muscle-type (M). Five possible combinations of subunits (isoenzymes) exist: H_4, H_3M_1, H_2M_2, H_1M_3, M_4. The five isoenzymes have different kinetic characteristics. Different tissues have characteristic ratios of the five isoenzymes that suit particular metabolic needs.

1.3.7 Enzymes that lack a quaternary structure

Enzymes that lack a quaternary structure are known as **monomeric** enzymes, comprising a single polypeptide chain. Very few monomeric

enzymes are known and all of these catalyse a hydrolytic reaction. Monomeric enzymes typically comprise 100 to 300 amino acid residues and have molecular weights in the range 13 000 to 35 000.

Many proteolytic enzymes (proteases) are monomeric enzymes. These enzymes catalyse the hydrolysis of peptide bonds in other proteins and may be active against a variety of protein molecules. They are usually biosynthesised as larger inactive precursors known as **zymogens** (or **proenzymes**) to prevent generalised damage to cellular proteins. Zymogens are activated as required by the action of other proteolytic enzymes (also see Section 1.7).

1.4 Forces that stabilise protein molecules

The conformation of a protein is ultimately defined by its amino acid sequence, i.e. its primary structure. Upon biosynthesis of a polypeptide, folding is thought to occur simultaneously in many places, giving rise to secondary structure. The folded regions then interact to form structural motifs and domains and, finally, the domains associate to give rise to the tertiary structure of the protein, i.e. its native conformation. Whereas regions of secondary structure are stabilised by interactions between amino acids close together within the polypeptide chain, tertiary structure is stabilised by interactions between amino acids that are far apart on the chain but that are brought into close proximity by protein folding. The main forces that stabilise tertiary structure are hydrophobic interactions, electrostatic interactions and covalent linkages. Table 1.2 compares bond strengths of interactions involved in stabilising protein molecules.

Hydrophobic interactions arise through the tendency of non-polar hydrophobic amino acid residues, such as leucine and valine, to minimise contact with water molecules. The most energetically favourable way for a protein to fold is therefore to enclose the hydrophobic side chains of such amino acids in the interior of the protein, i.e. remove them from the

Table 1.2 Comparative bond strengths of interactions involved in stabilising protein molecules

Interaction type	Approximate bond energy ($kJ\,mol^{-1}$)
Disulfide bond	200
Ionic interactions	86
Hydrogen bond	10–20
Van der Waals forces	10
Hydrophobic interactions	0

surrounding aqueous (polar) environment. Hydrophobic interactions are the most important stabilising influence on protein tertiary structure.

Electrostatic interactions include ionic interactions, hydrogen bonds and Van der Waals forces. Ionic interactions (salt bridges) occur between oppositely charged amino acid side chains, e.g. between the amino group (—NH_3^+) of lysine and the carboxyl group (—COO^-) of glutamate. Ionic interactions, which are found mainly on the surface of folded polypeptides, contribute only modestly to tertiary structure. The strength of interaction is determined by the proximity of the groups and the medium that separates them, the force between the two charges being inversely proportional to the dielectric constant of the medium. Hydrogen bonds are stronger than hydrophobic interactions and occur extensively in proteins (Table 1.2). However, they do not contribute substantially to overall conformational stability of a polypeptide because in an unfolded state, hydrogen bonds would form with surrounding water molecules, i.e. there is little energy difference between folded and unfolded states. These bonds, however, do influence protein folding and, therefore, the native tertiary conformation, because they are strongest when the atoms involved are arranged linearly, e.g. in an anti-parallel β-sheet (Fig. 1.7c).

Van der Waals forces are relatively weak electrostatic interactions that occur between atoms and molecules when they are an optimum distance apart (Table 1.2). Electrons (negatively charged) of an atom become attracted to the nuclei (positively charged) of other atoms, giving rise to dipoles. Although each van der Waals interaction is weak, there are large numbers of interactions and overall they contribute significantly to tertiary protein structure.

Covalent bonds are far stronger than electrostatic interactions. Disulfide linkages are the major covalent bonds that help stabilise the tertiary structure of a protein and occur between cysteine residues. They are formed after the protein has folded into its native tertiary conformation and therefore stabilise the structure but do not direct folding. Disulfide bonds are rarely formed in intracellular proteins because of the reducing environment. They are formed in proteins exposed to more oxidising environments such as many extracellular proteins and outer membrane proteins. Disruption of disulfide bonds often has little effect on the tertiary structure, although in cysteine-rich polypeptides the protein tends to become less conformationally stable.

Polypeptide chains of oligomeric proteins are usually linked together by non-covalent interactions; peptide bonds are never involved.

1.5 Properties of proteins

The chemical properties of proteins are essentially those of the side chains of the amino acid residues. Several chemical reactions with proteins

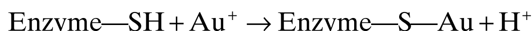
Table 1.3 Some of the reactions that can be used to estimate the quantity of protein in solution

Reaction name	Reactants	Colour produced
Hopkins–Cole	Indole group of tryptophane / glyoxylic acid and sulfuric acid	Purple
Millon	Phenolic group of tyrosine / mercuric sulfate and sodium nitrate	Red
Folin–Ciocalteu	Phenolic group of tyrosine / tungstate and molybdate	Blue
Sakaguchi	Guanidine group of arginine / α-naphthol and oxidising agent	Red
Biuret	Peptide bonds / alkali cupric sulfate	Purple
Lowry	Combination of biuret and Folin–Ciocalteu reactions	Blue

produce coloured products, where the intensity of the colour produced is proportional to the number of reacting groups present (Table 1.3). Such reactions can be used for quantitative estimation of protein concentration in solutions. For many of these reactions it is necessary to assume that the protein being estimated has an average distribution of amino acid residues and that free amino acids (or other reacting groups) are not present. In the case of the Biuret reaction (and Lowry method), however, free amino acids do not react because it is based on reaction with the peptide bonds.

Although most proteins are not coloured, they all absorb light in the ultraviolet (UV) region of the spectrum. UV absorbance of proteins is particularly strong at 280 nm owing to the side chains of tryptophan and tyrosine. Measurement of protein absorbance at this wavelength can be used as a ready but rough estimation of protein in solution.

Functional groups in amino acid side chains have an important role in enzymatic catalysis. This is considered later in this chapter (Section 1.8) and in Chapter 3. Inactivation of enzymes can also occur by binding of chemical agents to these functional groups. For example, sulfydryl groups of cysteine bind heavy metal ions strongly, which can inactivate many enzymes:



$$\text{Enzyme—SH} + \text{Au}^+ \rightarrow \text{Enzyme—S—Au} + \text{H}^+$$

1.5.1 Acid–base properties

Proteins contain ionisable groups and therefore display acid–base properties. By definition, an acid is a proton donor and a base is a proton acceptor. Any acid–base reaction may be written as:

$$\text{HA} \Leftrightarrow \text{H}^+ + \text{A}^-$$

where HA is an acid and A^- is a base.

The strength of an acid depends on its dissociation constant, K_a. A strong acid dissociates readily, whereas a weak acid dissociates only partly (i.e. has a higher affinity for its proton). Generally, dissociation (K_a) of an acid may be described as:

$$K_a = \frac{[H^+][A^-]}{HA}$$

Rearranging:

$$[H^+] = \frac{K_a[HA]}{[A^-]}$$

Taking the negative logarithm of both sides:

$$-\log_{10}[H+] = -\log_{10} K_a + \log_{10}\frac{[A^-]}{[HA]}$$

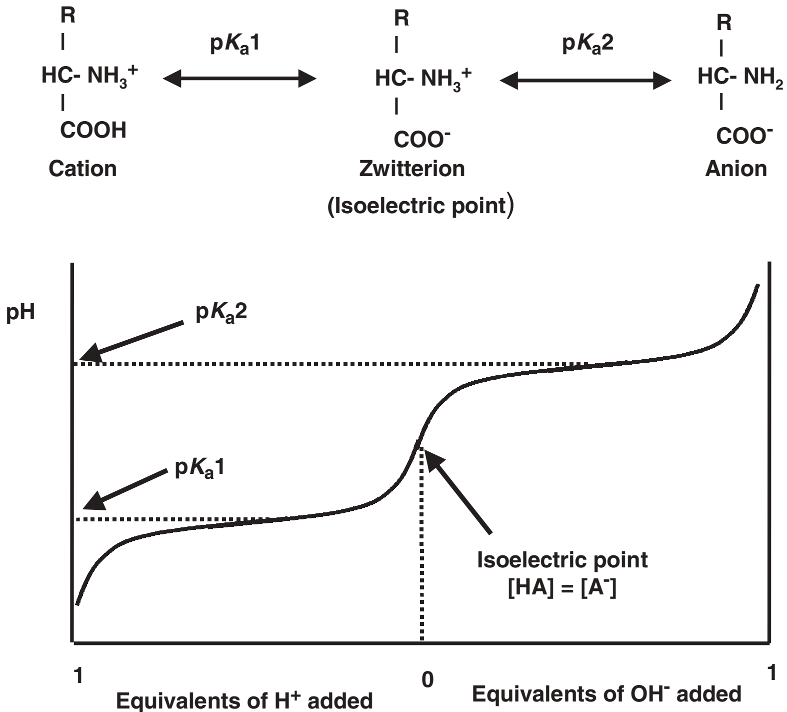
Defining pH and pK_a:

$$pH = pK_a + \log_{10}\frac{[A^-]}{[HA]}$$

The final expression gives the relationship between pH and the degree of ionisation of the ionisable species [HA]. It is known as the Henderson–Hasselbalch equation.

All amino acids will be ionised in aqueous solution in a manner determined by pH. From Fig. 1.10 it can be seen that the pK_a values are the points of inflection on the curve. Two ionisation curves are shown, giving two pK_a values, because the α-carboxyl and α-amino groups of the free amino acid are ionised over different pH ranges. If several ionisable groups are present, as is the case for an enzyme, the situation is more complicated. In proteins, the ionisable groups contributing to the acid–base properties are the C-terminal α-carboxyl group, the N-terminal α-amino group and the side chains of certain other amino acids (arginine, aspartate, cysteine, glutamate, histidine, lysine and tyrosine). Other α-carboxyl and α-amino groups do not contribute to acid–base properties because they form the peptide bonds in the protein. Approximate pK_a values range from 3.0 for the α-carboxyl group, through 8.0 for the α-amino group to 12.5 for the guanidine group of arginine. The overall titration curve for a protein will comprise the effects of all ionisable groups present in the molecule.

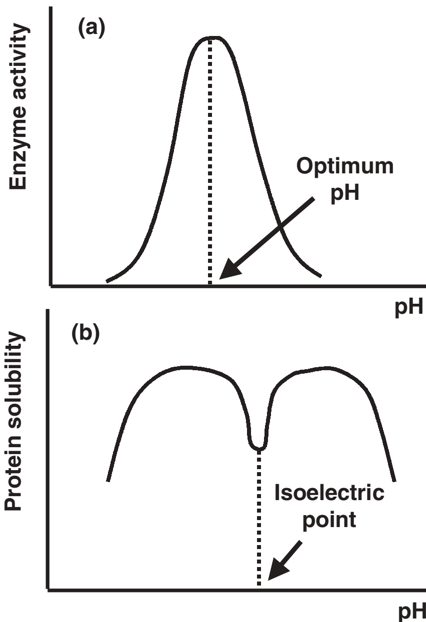
The titration curve for an amino acid shown in Fig. 1.10 shows that the buffering capacity of an ionisable group is greatest where the pH is near its pK_a value. Globular (soluble) proteins act as buffers that have the greatest buffering capacity near the pK_a value of the most frequently occurring ionisable group in the molecule.



1.10 Titration curve for an amino acid.

Just as the charge on a free amino acid is determined by pH, as shown in Fig. 1.10, so is that on a protein. At low pH most proteins carry a net positive charge, whereas at high pH most proteins carry a net negative charge. The pH at which there is an equal balance between negative and positive charge (i.e. no net charge on the protein) is known as the isoelectric point.

Enzyme functioning is pH dependent and each enzyme has a characteristic pH at which catalytic activity is highest (optimum pH), as well as minimum and maximum values beyond which the enzyme does not function. Most enzymes are active within the pH range 5 to 9 and many display a bell-shaped curve of activity against pH, as illustrated in Fig. 1.11a. The influence of pH on enzyme activity is a result of a combination of (1) the ionisation state of the amino acid side chains involved in the catalytic activity of the enzymes and (2) the ionisation of the substrate. Binding of substrate to the enzyme lysozyme, for example, is related to the ionisable state of two key amino acids: asparagine at position 52 and glutamate at position 35. The enzyme catalyses hydrolysis of the major cell wall polyscaccharides of bacterial cells. At the pH optimum for lysozyme (pH 5), asp-52 is ionised ($—COO^-$), while glu-35 is not (—COOH). The ionisable state of these two

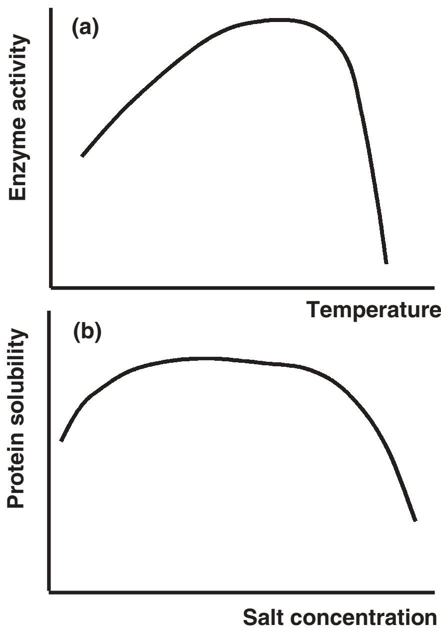


1.11 Typical profiles for influence of pH on (a) enzyme activity and (b) protein solubility.

amino acids has a strong influence on substrate binding and thus enzyme activity. Activity decreases rapidly below the pH optimum because asp-52 becomes protonated (—COOH). At pH values above the optimum, activity decreases because glu-35 becomes ionised ($—COO^-$). At extremes of pH, the charge on the ionisable side chains causes the compact tertiary structure (see Section 1.3.5) to disrupt and the protein denatures. For soluble (globular) proteins, denaturation leads to a marked reduction in solubility since hydrophobic groups normally on the inside of the molecule become exposed to the aqueous solvent. The solubility of an enzyme also decreases around its isoelectric point (Fig. 1.11b). The mechanism for this is formation of insoluble aggregates of the neutral molecules at the isoelectric point which, at other pH values, would have been repelled by the net charge they carry.

1.5.2 Other factors influencing protein solubility

Temperature and **salt concentration** are other important factors influencing the solubility of enzymes in aqueous solution. Enzyme activity increases with temperature up to an optimum temperature beyond which activity decreases sharply as the tertiary structure is disrupted through thermal



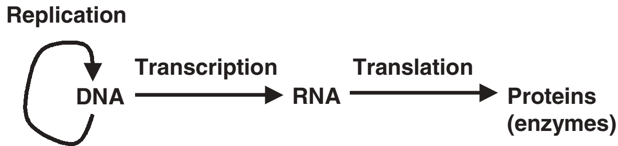
1.12 Typical profile for influence of (a) temperature on enzyme activity and (b) salt concentration on protein solubility.

agitation leading to denaturation (Fig. 1.12a). Change in solubility of enzymes with temperature largely mirrors that for change in activity.

At low salt concentration, the solubility of an enzyme increases owing to changes in ionisation of amino acid side chains. At very high salt concentrations, however, the high level of interaction between the added ions and aqueous solvent reduces the interaction between the protein and water, leading to precipitation of the enzyme from solution (Fig. 1.12b).

1.6 Biosynthesis of proteins

All cells have deoxyribose nucleic acid (DNA) as their genetic material, which ultimately codes for all the components of the cell including its enzymes. There are two essential processes that convert the information in DNA into protein. Transcription produces a ribose nucleic acid (RNA) copy, or messenger molecule, which is then translated into protein. These processes form the so-called central dogma of modern biology (proposed by Crick in 1956), which proposes a one-way flow of genetic information from DNA, through RNA to protein:



i.e. DNA makes RNA makes protein

The central dogma remains essentially unchanged today, although some rare exceptions to the rule of one-way flow of information from DNA to RNA are known. A detailed consideration of the central dogma is beyond the scope of this chapter. This section will explain, in broad terms, the process by which the sequence of nucleotides in RNA is converted into a sequence of amino acids in proteins – the process termed **translation**. A basic knowledge of the structure of DNA including its base pairing (i.e. A-T and C-G) is assumed. As the central dogma indicates, the base sequence of DNA indirectly specifies the amino acid sequence in proteins, i.e. an intermediate RNA molecule is involved that directly specifies the primary polypeptide structure. Other genes in DNA code for RNA molecules with a different function (see below). The enzymes that transcribe DNA to RNA are called RNA polymerases. Unlike DNA, RNA is usually a single strand of polynucleotide because it is formed by copying the base sequence of only one of the strands of DNA. In RNA, uracil replaces the thymine base in DNA. The RNA code therefore comprises adenine (A), uracil (U), guanine (G) and cytosine (C). Three major types of RNA are involved in protein synthesis: messenger RNA (mRNA), transfer RNA (tRNA) and ribosomal RNA (rRNA).

For protein synthesis to occur, the base sequence of DNA must be transcribed by RNA polymerase II to produce a mRNA molecule (the process of translation). It is the base sequence of the mRNA that codes directly for the amino acid sequence of the protein. Each amino acid is coded for by a triplet of bases (known as a codon). Since there are four different bases in RNA, there are 4^3 (=64) different triplet codons. It follows that since there are only 20 amino acids found in proteins, many must be coded for by more than one codon. The sequence of nucleotides in a codon that specifies a given amino acid is known as the genetic code and is shown in Fig. 1.13. Because more than one codon can determine the same amino acid the code is said to be 'degenerative'. Where degeneracy occurs it tends to occur in the third nucleotide of the codon with U and C and also A and G being equivalent to each other. Serine, for example, can be coded for by UCU, UCC, UCA or UCG. Degeneracy has the advantage that a mutation held in the third nucleotide has a lower chance of modifying the code than a

Second nucleotide position

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

t.c. = termination codon

1.13 Genetic code: the sequence of nucleotides in a codon that specifies a given amino acid.

change in the other two nucleotides of the codon. As shown in Fig. 1.13, the code is not ambiguous as no codon codes for more than one amino acid.

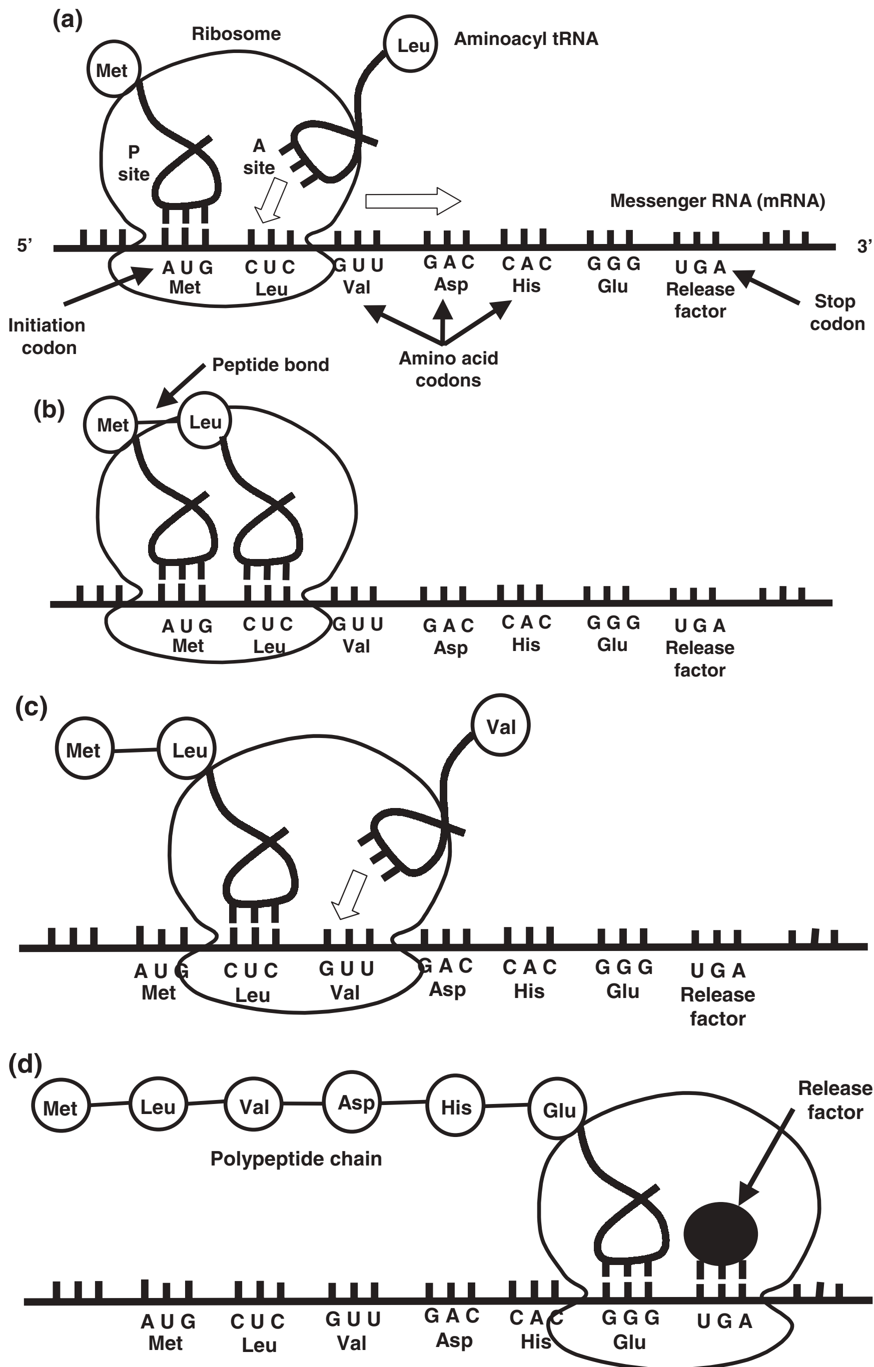
Codon triplets of nucleotide bases cannot themselves recognise amino acid molecules. For protein synthesis, adaptor molecules are therefore required to align up amino acids on adjacent mRNA codons. The adaptors are tRNA molecules that are linked to their particular amino acid by highly specific enzymes (amino-acyl tRNA synthetases). There are at least 20 different types of these enzymes per cell: one for each amino acid. tRNA molecules are transcribed from DNA by RNA polymerase III enzymes. Each tRNA has a complementary sequence of bases (known as an anticodon) for the codon for its amino acid, which recognises the correct codon on the mRNA. The amino acids of the poypeptide are therefore arranged in the sequence defined by the order of codons in the mRNA. Linking an amino acid to its tRNA has the effect of activating it for participation in protein synthesis. The high specificity of the enzyme for the amino acid and the tRNA to which it is joined is essential for reliable protein synthesis. Subsequent translation of the mRNA code to give the polypeptide occurs on

particles in the cytoplasm called ribosomes. These particles are composed of proteins and rRNA. Ribosomes from both prokaryotic and eukaryotic cells are made up of large and small subunits. They have two grooves: one accommodates around 35 nucleotides of the mRNA molecule, the other accommodates the growing polypeptide chain. Ribosomes also have two binding sites for tRNA molecules, one for peptide (termed P) and one for amino acid (termed A). The orientation of the P and A sites are such that for a tRNA molecule to bind, its anticodon must be complementary to the codon on the mRNA.

Figure 1.14 illustrates the process of translation of the message of mRNA into a polypeptide chain or protein. Once the P-site is occupied, an aminoacyl tRNA binds to the vacant A-site. The peptide chain grows by the carboxyl end becoming detached from the P-site tRNA and forming a peptide bond with the amino end of the amino acid at the A-site tRNA. The tRNA then leaves the P-site and the peptide-tRNA on the A-site moves to occupy the vacant site. It is the ribosome that moves in this process, not the tRNA. The movement involves a conformational change in one of the ribosomal proteins and is an energy-requiring step. Once the A-site is vacant the process starts again, involving binding of another aminoacyl tRNA. Each cycle takes only a twentieth of a second and bacteria can synthesise protein containing 300 amino acid residues in about 20 seconds.

Protein synthesis (translation) has three main stages: initiation, elongation and termination. Translation in eukaryotic cells is initiated by a special tRNA (initiator tRNA) recognising the start code of protein synthesis; the codon AUG. The initiator tRNA carries a methionine residue (tRNA-met) and is specific for the start codon. The initiator binds with the small ribosome subunit, in a process also involving initiation factors, to form a complex (Fig. 1.14a). The tRNA-met-small subunit-initiation factor complex associates with a mRNA molecule. Binding of tRNA-met with the start codon AUG then occurs and the initiation factors are released. Only then does the large ribosomal subunit bind to form the complete ribosome, with tRNA-met on the P-site. Aminoacyl tRNA can then bind to the A-site and protein synthesis proceeds (Fig. 1.14a). The initiation process so described is similar in prokaryotic cells, except that the initiator tRNA is linked to *N*-formyl methionine and not methionine.

In the elongation stage of protein synthesis, the formation of the peptide bond is catalysed by the enzyme peptidyl transferase (Fig. 1.14b). Addition of amino acids to the growing peptide chain continues until a stop codon in the mRNA is in the A-site of the ribosome (Fig. 1.14c,d). This leads to termination, in which a protein release factor binds to the stop codon and causes peptidyl transferase to hydrolyse the peptide-tRNA link (i.e. water molecule added instead of an amino acid). The polypeptide moves away from the ribosome and spontaneously coils into its secondary and tertiary



1.14 Protein biosynthesis in eukaryotes. (a) Initiation of protein biosynthesis involving binding of tRNA-met to the P-site and of an aminoacyl tRNA to the A-site. (b) Peptide bond formation catalysed by peptidyl transferase. (c) Elongation of the polypeptide chain. (d) Termination of protein synthesis involving binding of a protein release factor.

structure. The functional ribosome then disassembles, leading to separation of small and large subunits, tRNA and release factor.

The addition of an amino acid to a growing polypeptide chain is energy demanding, requiring four high-energy bonds to make a new peptide bond; protein synthesis is the most energy demanding of all cellular biosynthetic processes.

1.7 Post-translational modification of proteins

Post-translational modification refers to the covalent modification of many proteins that occurs during or after their formation on ribosomes. Phosphorylation, glycosylation and proteolytic processing are the most common forms of post-translational modification, although several others are known to occur. These modifications tend to affect the structural stability or biological activity of the protein.

Reversible **phosphorylation** of proteins occurs through the action of specific protein kinase and protein phosphatase enzymes, giving rise to phosphorylated and dephosphorylated forms of the protein, respectively. Adenosine triphosphate (ATP) is the usual phosphate group donor. The phosphate group is transferred to the hydroxyl group of serine, threonine or tyrosine residues, although the side chain of certain other amino acids can also be phosphorylated. In most cases, phosphorylation–dephosphorylation functions as a reversible on-off switch with respect to biological activity. Some proteins are biologically activated by phosphorylation, whereas others are inactivated. Phosphorylation also alters the physiochemical properties of a protein through increase in the number of negative charges carried.

Glycosylation refers to the attachment of carbohydrates. It is a common post-translational modification of extracellular and cell surface proteins of eukaryotic cell origin. There are two types of glycosylation: N-linked and O-linked. In N-linked glycosylation the oligosaccharide (carbohydrate chain) is linked to the nitrogen atom of an asparagine residue, whereas O-linked glycosylation involves oligosaccharide linkage to the oxygen atom of hydroxyl groups (usually of serine or threonine residues). The oligosaccharide component of a glycoprotein has various potential functions: direct role in the biological activity of the protein; targeting of proteins to specific cellular locations; helping stabilise the protein; protecting from proteolytic attack; enhancing solubilisation of protein; increasing biological half-life of protein; enhancing recognition and cell–cell adhesion.

Proteolytic post-translational modification occurs for some proteins. Generally, these proteins are inactive when synthesised (the ‘pro’ form or precleaved state) with functional activation occurring upon proteolysis; also see Section 1.3.7. Such proteolytic cleavage is very specific and generally

irreversible. Most of the proteins that are processed in this way are ultimately exported to cellular organelles or secreted from the cell.

Acetylation (CH_3CO—group transfer) of proteins can occur during synthesis or post-translationally. The enzyme involved is *N*-acetyl transferase, which is loosely associated with ribosomes and usually uses acetyl-CoA as acetyl donor. More than half of all polypeptides synthesised in the cytoplasm have an N-terminal acetyl group. The function of N-terminal acetylation and the criteria governing selection of polypeptides for this modification are not fully understood.

Protein **acylation** is thought to occur in all eukaryotic cells, and a wide range of cytoplasmic and membrane proteins are modified through acylation. In this process, polypeptides are modified by direct covalent attachment of fatty acids. The saturated fatty acids, palmitic acid (C_{16}) and myristic acid (C_{14}) are most commonly involved. Palmitic acid is attached through ester or thioester bond formation to either a cysteine, serine or threonine residue. Several enzymes are thought to be involved in this type of post-translational protein modification. Myristic acid is always covalently attached to proteins via an amide bond to an N-terminal glycine residue. This modification occurs before polypeptide synthesis is complete and is catalysed by the enzyme myristoyl CoA:protein *N*-myristoyl transferase. Acylation is thought to play a role in interaction of proteins with biological membranes or with other proteins. Many acylated polypeptides are found in multi-subunit protein complexes and removal of the fatty acid component has a negative affect on subunit interaction.

1.8 Enzymatic catalysis

Enzymes are catalysts – they speed up the rate of chemical reactions but remain unaltered themselves.

Chemical reactions proceed when the free energy of the products is less than that of the reactants. In spontaneous reactions therefore, the change in free energy (ΔG) is positive, i.e. energy is released. Although biochemical reactions of cellular metabolism are spontaneous in the sense that they release energy, they proceed exceedingly slowly in the absence of enzymes because of the energy 'barrier' between substrate(s) and product(s). This barrier is known as the 'activation energy' and for molecules to react, they must possess energy corresponding to the top of this barrier. Heating a solution containing substrate molecules is one way of increasing the energy levels. Enzymes, however, do not increase energy levels of substrate molecules; rather they provide an alternative low-energy route for the reaction to proceed. Enzymes therefore function as catalysts by lowering the activation energy of a reaction. The effect is to accelerate spontaneous (energy releasing) reactions. Essentially, more molecules of substrate have sufficient

energy to react in an enzyme-catalysed reaction. Some enzymes have enormous catalytic power. For example, catalase which catalyses the breakdown of hydrogen peroxide ($2H_2O_2 \rightarrow 2H_2O + O_2$) accelerates the reaction around 10^{14} times (a million billion) compared to the uncatalysed reaction, each molecule of catalase being capable of processing around five million molecules of hydrogen peroxide per second!

Enzyme catalysis involves formation of a transition state (ES) in which the substrate is bound to the enzyme, i.e. $E + S \leftrightarrow ES \rightarrow E + P$. Generally, anything that promotes the formation of the transition state will enhance the rate of reaction.

Mechanisms of enzyme catalysis are considered in Chapter 3.

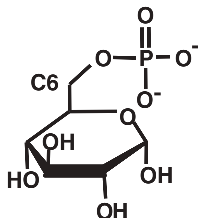
1.8.1 Specificity

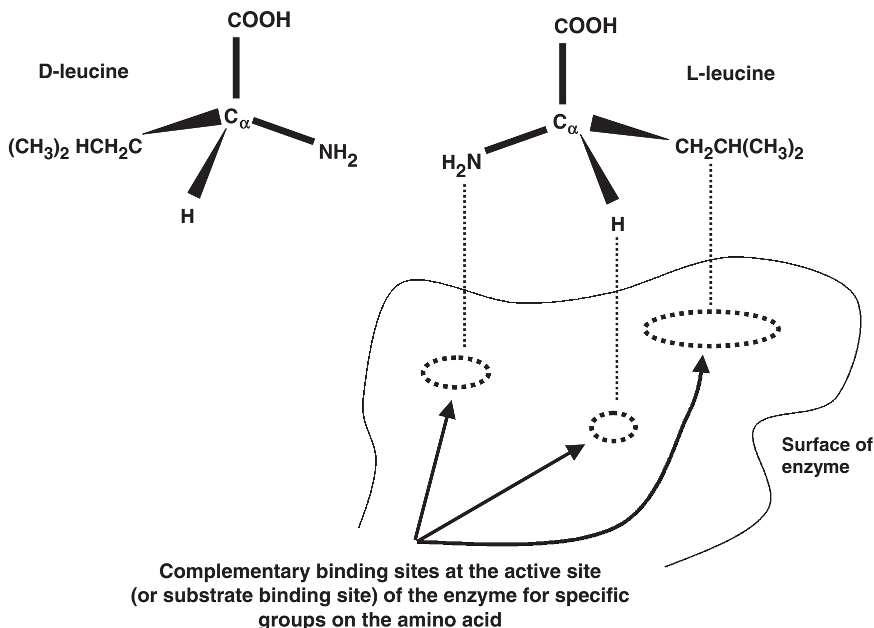
Enzyme specificity depends on two important aspects:

- What molecule(s) any particular enzyme acts on;
- What reaction is performed.

All enzymes can discriminate between different molecules but the extent of discrimination varies between different enzymes. The enzyme urease, for example, is highly specific, being capable of acting only on urea (H_2NCONH_2) and not on closely related molecules (e.g. $H_2NCONHCH_3$). Other enzymes can act on closely related molecules. The proteolytic enzyme trypsin, for example, hydrolyses the peptide bond to the C-terminal side of either arginine or lysine. Trypsin is most active at pH 8.0, at which arginine or lysine are the only amino acids with side chains that are positively charged. The enzyme therefore tolerates minor differences in the shape of its substrate, provided that the amino acid side chain is positively charged.

Most enzymes are highly specific with regard to which groups and bonds are acted upon, and also therefore what products are formed. This positional or regional specificity (regiospecificity) is vital to cellular metabolism. The enzyme hexokinase that catalyses the initial phosphorylation of glucose in metabolism displays absolute positional specificity; although glucose has five hydroxyl groups available for phosphorylation, the enzyme always adds the phosphate group at the same C6 position.



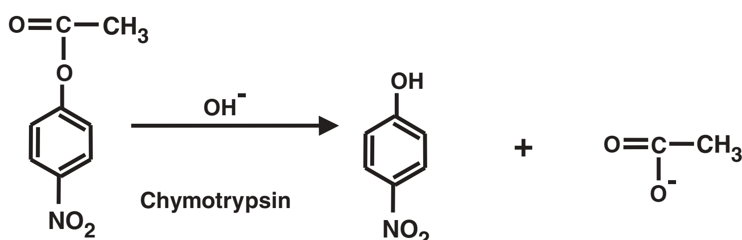


1.15 Enzyme discrimination between stereoisomers.

Enzymes also display high specificity with regards to stereoisomers. Enzymes that act on optically active compounds normally show absolute specificity for either the D- or the L-isomer. For example, hexokinase is active on D-glucose and not on L-glucose, whereas propteolytic enzymes are active towards polypeptides made up exclusively of L-amino acids. As indicated earlier in this chapter (Section 1.3.2), sterioisomers arise in organic molecules when an asymmetric centre occurs, i.e. there is a carbon bonded to four different atoms or groups. The only difference between the pairs of isomers is the distribution of the substituent atoms or groups around the centre carbon atom. In free solution these isomers will react in the same manner. However, when bounded to an enzyme with complementary sites for three of these groups, the position of the binding sites dictates which of the two isomers binds and then reacts. This is illustrated in Fig. 1.15 for an amino acid leucine, where only L-leucine binds to the enzyme. Although binding of substrate to the enzyme might confirm absolute stereospecificity, it might not necessarily be the case. Isomers would not be distinguished by an enzyme that required only two binding sites to be complementary, rather than three.

Some enzymes are far less specific and there are circumstances in which absolute specificity is not desirable to organisms. Hydrolytic enzymes, including many proteases, tend to display a relative lack of specificity. These

enzymes are often involved in breakdown of large molecules, releasing component parts that may be utilised as a growth nutrient. Chymotrypsin, for example, cleaves polypeptides that have an amino acid with an aromatic or other hydrophobic side chain on the carbonyl side of the peptide bond. This enzyme can also act on small ester substrates, such as *p*-nitrophenyl acetate:

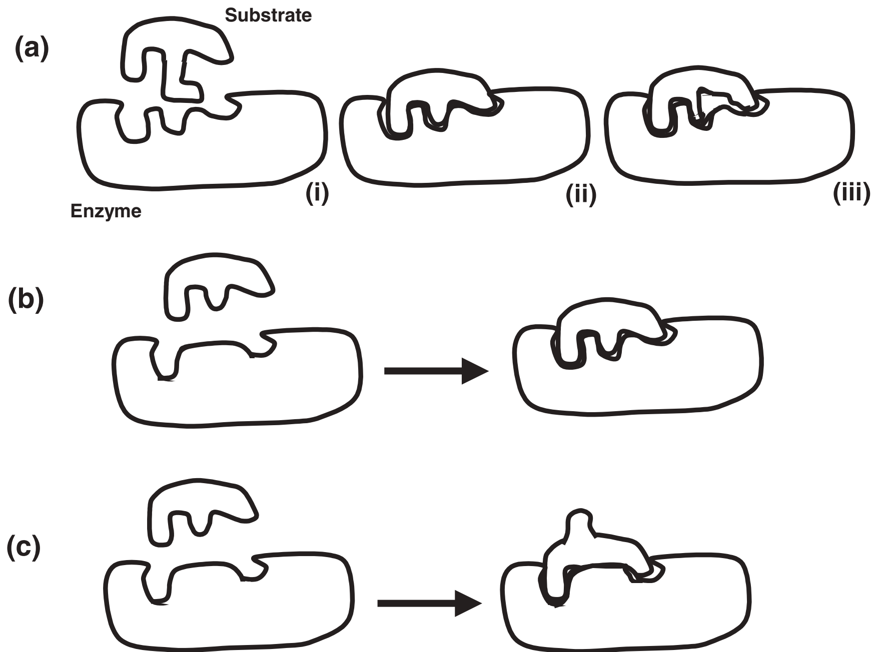


The enzyme, whether acting on *p*-nitrophenyl acetate or polypeptides, acts as an esterase.

Enzymes distinguish between different molecules by the particular shape and charge distribution of the active site of the enzyme. The active site is the region that binds the substrate and is only a small part of an enzyme. For some enzymes, as few as 5–6 amino acid residues within the protein form the active site. The active site must ensure not only that the substrate binds, but also that other molecules are excluded and that the substrate is in the correct orientation for catalysis to take place.

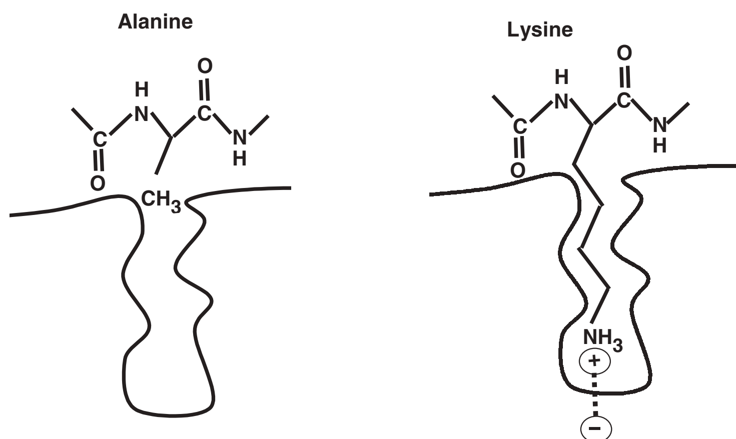
Fischer, as far back as 1890, proposed that the shape of the active site of an enzyme was complementary to that of its substrate and that this explained the specificity of enzyme action. This became known as the 'lock and key' model of enzyme–substrate interaction. The model provides a convenient explanation not only of the absolute specificity of some enzymes, but also of the more relaxed specificity of others (Fig. 1.16a). Enzymes, however, are now known to be flexible molecules and the lock and key model implies a rigid structure. Koshland, in 1958, proposed an alternative model for enzyme action in which the enzyme itself undergoes conformational change as a result of substrate binding (Fig. 1.16b). In this 'induced fit' model, the substrate induces a change in the shape of the active site to the correct fit. This occurs only when the substrate binds, which implies that the shape of the active site and its substrate are complementary only when in a transitional state of an enzyme–substrate complex. An enzyme may also distort (or strain) the substrate (Fig. 1.16c).

Enzymes bind to substrates by a combination of hydrogen and ionic bonds, as well as by hydrophobic and van der Waals interactions. Covalent



1.16 Models of enzyme active site interaction with substrate. (a) The 'lock and key' model for an enzyme active site with relaxed specificity: (i) no binding of incompatible substrate; (ii) binding of substrate having high complementarity with active site; (iii) binding of substrate having low complementarity with active site. (b) The 'induced fit' model in which the substrate induces a change in the shape of the active site to the correct fit. (c) Deformation (or straining) of substrate leading to improved fit with active site.

bonds are only occasionally formed and then usually as part of the catalyst. The strength of binding is a function of the distance apart of atoms; repulsion occurs if too close and increasing distance results in a progressively weaker interaction. Generally, strong binding results from complementarity of both shape and charge for the active site and its substrate. This is illustrated in Fig. 1.17 for a proteolytic enzyme capable of binding both alanine and lysine. The active site makes very limited contact with the side chain of alanine, resulting in low strength binding and a low tendency to bind. For lysine, however, the side chain makes numerous close contacts (van der Waals interactions) with the active site and an ionic bond is formed at the bottom of the active site; both the strength and tendency of lysine binding are relatively high.



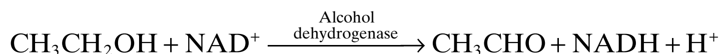
1.17 Complementarity of substrate and active site of an enzyme governs strength of binding.

1.8.2 Enzyme cofactors

The catalytic activity of some enzymes requires the presence of non-proteinaceous compounds. Such compounds are collectively known as cofactors. Enzymes that require cofactors are often commonly referred to **apoenzymes** when not bound to the cofactor and are catalytically inactive, and as **holoenzymes** when bound to the cofactor and catalytically active (Fig. 1.18).

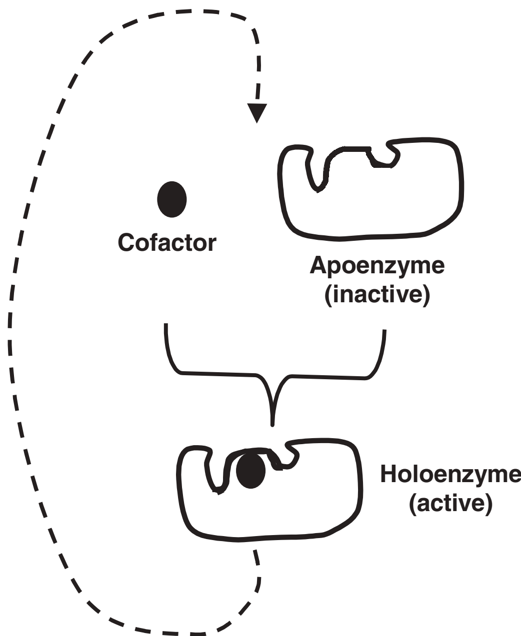
Cofactors are broadly categorised into three types:

1. Prosthetic groups are organic cofactors that are tightly bound to the enzyme, sometimes covalently. Flavin adenine dinucleotide (FAD) of some dehydrogenase enzymes and biotin of some carboxylase enzymes are examples of prosthetic groups (Fig. 1.19).
2. Coenzymes are organic cofactors that are more easily removed from the enzyme than prosthetic groups. Examples are nicotinamine adenine dinucleotide (NAD^+) or nicotinamine adenine dinucleotide phosphate ($NADP^+$), which are required by many dehydrogenase enzymes for catalytic activity (Fig. 1.19). NAD^+ and $NADP^+$ accept reducing equivalents from a variety of substrates depending on the particular enzyme. For example:



$$CH_3CH_2OH + NAD^+ \xrightarrow[\text{dehydrogenase}]{\text{Alcohol}} CH_3CHO + NADH + H^+$$

Coenzymes can often be regarded as a second substrate for the enzyme. The dehydrogenase enzymes, for example, have a strong binding site for the oxidised form of their coenzyme (NAD^+). Once the substrate is



1.18 Cofactor interaction with an enzyme. Some cofactors (mostly coenzymes) are not bound tightly to enzymes and are readily lost when enzymes are extracted from cells, leading to inactive forms of the enzymes. Enzyme activity is normally restored by addition of the essential cofactor.

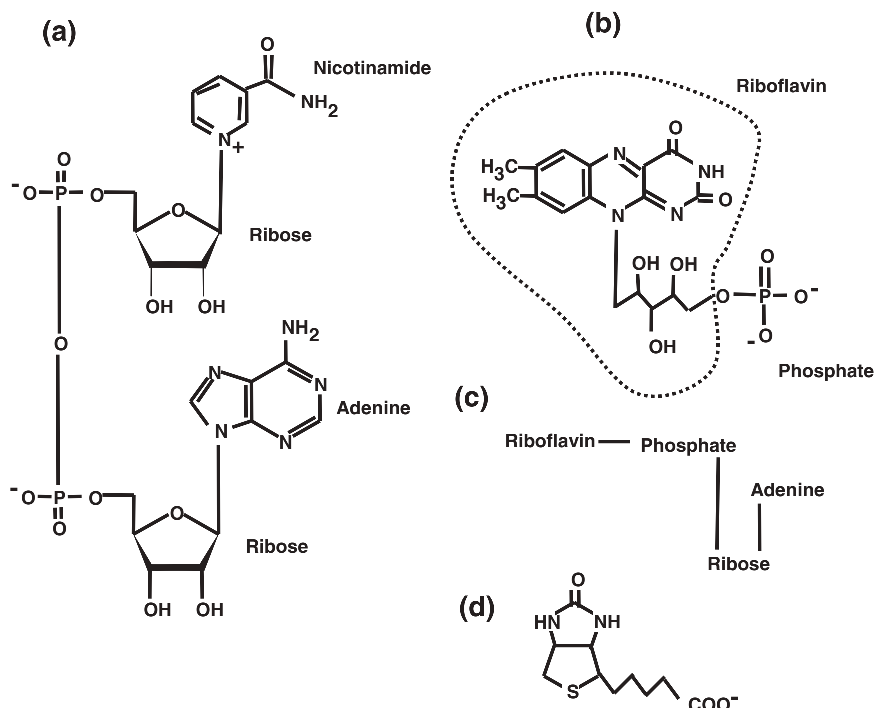
oxidised, the reduced form (NADH) leaves the enzyme and is reoxidised in cellular metabolism. The NAD^+ so formed is available for binding to the dehydrogenase enzyme and the cycle is repeated. In this way, NAD^+ acts as a second substrate for the enzyme, but unlike most substrates it must be continually recycled within the cell.

3. Metal ions are cofactors for some enzymes. They may be loosely or tightly bound and, in some cases, are associated with the prosthetic group (e.g. Fe^{2+} or Fe^{3+} in the heme group). These ions are similar to coenzymes in the sense that they confer on the enzyme a property it would not possess in their absence. In other cases, free metal cations such as Fe^{3+}, Zn^{2+}, Cu^{2+} and Mg^{2+} function as cofactors.

Table 1.4 gives examples of cofactors and their catalytic roles.

1.9 Future trends

Technical advances in NMR and computational analysis of spectra are likely to enable structural determination of relatively high molecular weight



1.19 Structures of some important cofactors. (a) Nicotinamide adenine dinucleotide (NAD^+). $NADP^+$ differs from NAD^+ only by the presence of a phosphate group on the adenine. (b) Flavin mononucleotide (FMN). (c) Flavin adenine dinucleotide (FAD). (d) Biotin.

proteins in solution. Fourier Transform Infrared Spectrometry (FTIR) is also likely to contribute strongly to analysis of protein secondary structure in the future. The advantage of FTIR over other techniques is that spectra can be obtained for proteins in a wide range of environments, for example in solution, on various surfaces. Improved characterisation of infrared spectra for proteins with defined secondary structure content (using synthetic polypeptides), together with advances in computational analysis, is providing insights into the frequency and variability of particular secondary structural conformations. FTIR continues to offers great potential for investigating the influence of physical and chemical factors, such as temperature and pH, on the secondary structure of proteins in solution.

Through site-directed mutagenesis (or protein engineering) it is now possible to manipulate the amino acid sequence of a protein. The technique, which involves use of synthetic DNA fragments to change the genetic information (DNA) coding for a protein, an improved understanding of the

Table 1.4 Examples of enzyme cofactors and their roles

Cofactor	Role
Nicotinamide adenine dinucleotide (NAD^+, $NADP^+$)	Oxidation/reduction reactions [those involving transfer of H^- (hydride) ion]
Flavin adenine dinucleotide (FAD)	Oxidation/reduction reactions
Flavin mononucleotides (FMN)	Oxidation/reduction reactions
Biotin	Carboxyl group transfer
Cobalamin	Methyl group transfer
Coenzyme A (CoA)	Transfer of groups, e.g. acetyl
Thiamine pyrophosphate (TPP)	Acetaldehyde transfer
Tetrahydrofolate (THF)	One-carbon transfer reactions
Pyridoxal phosphate	Transamination and decarboxylation reactions
Metal ions:	
Fe, Cu, Mo	Oxidation/reduction reactions
Zn	Helps bind NAD^+
Co	Part of cobalamin coenzyme
Mn	Aids in catalysis by electron withdrawal

relationship between the primary amino acid sequence and its higher structural conformation can be obtained. Site-directed mutagenesis is also a valuable tool for studying structural and functional relationships of enzymes and for enhancing the attributes for commercially important enzymes. The catalytic capabilities of several enzymes have been improved by changes to the amino acids forming the active sites. Other studies have shown that it is possible to improve temperature and pH stability of enzymes. For example, introduction of cysteine residues into the polypeptide chain tends to enhance protein stability through disulfide bond formation. Replacement of lysine with arginine in a polypeptide chain also tends to enhance enzyme stability through increase in the extent of overall hydrogen bonding. Ultimately, continuing advances in this field could facilitate routine *de-novo* design of enzymes to suite particular applications; the 'Holy Grail' of applied enzymology!

The advances made in molecular biology and the associated research equipment have accelerated DNA sequence determination and large portions of the genome of several species have been sequenced. This genomic information will increasingly drive future trends in enzymology – within the emerging fields of **proteomics** and **bioinformatics** – leading to greatly improved understanding of structure, function and expression of newly discovered proteins.

1.10 Further reading

Modern textbooks of biochemistry[1–3] provide a wealth of fundamental information on enzymes and proteins in a well-illustrated and readily digestible form. Typically such texts have sections dealing with protein structure, enzymes as biological catalysts and protein expression in biological systems. They are an excellent source of background reading on these topics, and are recommended for improving understanding of fundamental biological aspects. These texts are also recommended for broader reading into modern molecular methods of DNA technology.

Two books[4,5] devoted solely to ‘enzymes’ are recommended for more in-depth studies. A notable strength of both texts is the simplified and effective approach to the complex mathematical treatment of enzyme kinetic data. There is also a useful survey[4] of relevant Internet sites and computer software for enzymatic data analysis. Palmer’s text[5] aims to provide a reasonably detailed account of all the various theoretical and applied aspects of enzymology likely to be included in a course on this subject – as a master educator, he delivers the material in a readily understandable form.

A book by Walsh[6] is also recommended as a source of information on industrial uses of proteins. It provides a broad overview of the various industrial uses of proteins, as well as an understanding of the fundamental biochemistry and methodological approaches that underpin such applications. The most notable feature that distinguishes it from standard molecular biology, biochemistry or biotechnology texts is the wide range of proteins considered and the depth of information provided. Sources of proteins are considered in some detail, including heterologous proteins from various microorganisms, plant and animal tissues and insect cells. In dealing with protein purification and characterisation, the theory is integrated well with the practical aspects and there is useful advice on scale-up of procedures. There is a chapter devoted to the various general issues relating to large-scale protein purification, with particular emphasis on the problems of contamination. Specific examples of how various proteins are produced commercially are also included in many of the subsequent chapters of the book. Chapters on industrial enzymes include – amongst others – proteases, carbohydrases, lignocellulases, pectinases, lipases and oxidoreductases. There is also an impressive collection of recent books and articles, with a good balance of general and specific sources of information.

A selection of review articles is provided[7–18] for those readers wishing to delve into a detailed consideration of particular aspects of protein and enzyme technology. These reviews are effective interfaces with the primary literature, which lead to key papers on the individual topics.

There are, of course, a host of Internet sites that provide information on proteins and enzymes, most of which provide useful links to other related sites. Structural computational analysis of proteins and detailed pictorial representations of proteins are notable strengths of the Internet in this context. A good starting point is The ExPASy (**Ex**pert **P**rotein **A**nalysis **Sy**stem) proteomics server of the Swiss Institute of Bioinformatics,[19] which is dedicated to the analysis of protein sequences and structures (see Reference section). A short list of other sites recommended of initial review is provided in the bibliography of this chapter.

1.11 Bibliography

1.11.1 Books

1. Garratt R H and Grishham C M (1999) *Biochemistry*, 2nd edition, New York, Saunders College Publishing.
2. Voet D and Voet J G (1999) *Fundamentals of Biochemistry*, Chichester, UK, John Wiley & Sons.
3. Stryer L (2002) *Biochemistry*, 5th edition, New York, W H Freeman.
4. Copeland R A (2000) *Enzymes: a practical introduction to structure, mechanism, and data analysis*, 2nd Edition, Chichester, UK, John Wiley & Sons.
5. Palmer T (2001) *Enzymes: biochemistry*, *biotechnology*, *clinical chemistry*, Chichester, UK, Horwood Publishing.
6. Walsh G (2002) *Proteins: biochemistry and biotechnology*, Chichester, UK, John Wiley & Sons.

1.11.2 Articles

7. Al-Lazikani B, Jung J, Xiang Z X and Honig B (2001) 'Protein structure prediction', *Curr. Opin. Chem. Biol.*, **5**, 51–56.
8. Arai M and Kuwajima K (2000) 'Role of molten globule state in protein folding', *Adv. Protein Chem.*, **53**, 209–282.
9. Chasse G A, Rodriguez A M, Mak M L, Deretey E, Perczel A, Sosa C P, Enriz R D and Csizmadia I G (2001) 'Pepide and protein folding', *J. Mol. Struct. – Theochem.*, **537**, 310–361.
10. Clore G and Gronenborn A (1998) 'NMR structure determination of proteins and protein complexes larger than 20 kda', *Curr. Opin. Chem. Biol.*, **2**, 564–570.
11. Harris J and Craik C (1998) 'Engineering enzyme specificity', *Curr. Opin. Chem. Biol.*, **2**, 127–132.
12. Merry T (1999) 'Current techniques in protein glycosylation analysis – a guide to their application', *Acta Biochem. Polon.*, **46**, 303–314.
13. Fagain C (1995) 'Understanding and increasing protein stability', *Biochim. Biophys. Acta*, **1252**, 1–14.
14. Parodi A (2000) 'Protein glycosylation and the affects of glycosylation on protein structure and activity', *Chembiochem*, **1**, 215–246.
15. Radford S (2000) 'Protein folding: progress made and promises ahead', *Trends Biochem. Sci.*, **25**, 611–618.

16. Robertson A and Murphy K (1997) ‘Protein structure and the energetics of protein stability’, *Chem. Rev.*, **97**, 1251–1267.
17. Schulein M (2000) ‘Protein engineering of cellulases’, *Biochim. Biophys. Acta – Protein. Struct. Mol. Enzymol.*, **1543**, 239–252.
18. Wilmouth R C, Clifton I J and Neutze R (2000) ‘Recent success in time-resolved protein crystallography’, *Natural Products Rep.*, **17**, 527–533.

1.11.3 Web sites

19. ExPASy (analysis of protein sequences and structures). *www.expasy.ch/*
20. Protein Data Bank (worldwide repository for the processing and distribution of 3-dimentional biological macromolecular structure data). *www.rcsb.org/pdb/*
21. Enzyme Nomenclature Database (repository of information relative to the nomenclature of enzymes primarily based on the recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology). *www.expasy.ch/enzyme/enzyme_details.*
22. SWISS-PROT (probably the most widely used protein database) on the site of European Bioinformatics Institure. *www.ebi.ac.uk/swissprot/*