8

Size Matters: Proteins and Enzymes

8.1 Principles of protein structure

Big molecules are special. Modern life has been transformed by large, man-made molecules, the synthetic polymers, such as plastics, fibres, adhesives, paints, and elastomers (rubbers). Their unique properties are related to the fact that the molecules are large. Small molecules simply cannot generate the same range of properties. Polymers are made up of smaller subunits, or monomers, which are joined by covalent bonds into long chains. The properties of the monomers will naturally influence the properties of the polymer, as will the number of each monomer in the polymer chain, because this will determine the final size of the polymer.

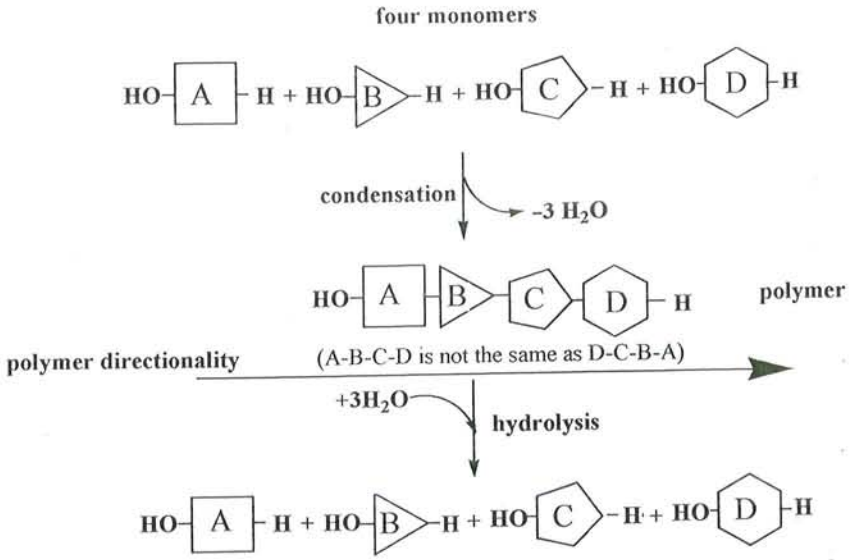
The ingenuity displayed by modern chemists to produce polymers with useful properties is impressive, and since the 1930s an enormous array of these useful materials has been produced. The end is not yet in sight. New polymers are being introduced on a regular basis because the possible combination of available monomers, plus the ways in which they may be arranged, is almost unlimited. Clever as it may be, all this endeavour is relatively insignificant compared with the products of biological evolution.

Big biological molecules have unique properties that allowed life on Earth to begin and to flourish. The ones that concern us most are the nucleic acids, the proteins, and the polysaccharides. These large biological molecules are built up in a living organism from smaller ones, and are best described as modular. All are made by the bonding of monomers-simple sugars to form polysaccharides, amino acids to form peptides and proteins, and nucleotides to form DNA and RNA. Although the monomers differ in chemical structure between the different types, there are features common to biopolymers worthy of mention:

- They are modular.
- They are formed biochemically by removal of water between the individual monomers, a process called dehydration or condensation. The condensation process is endothermic, using ATP as a source of energy.
- ATP requirements per monomer added: DNA/RNA 2, proteins 4, polysaccharides 2.
- They can be hydrolysed by enzymes back to the monomers.
- Each one has a characteristic three-dimensional architecture.
- Weak forces (secondary bonds) maintain their characteristic structure and help determine biomolecular interactions.

- Macromolecules and their monomers have a 'sense' or directionality, that is they do not 'read' the same in both directions.
- Macromolecules are informational, for example DNA and RNA.

Some of these properties are illustrated below.



Many monomers have molecular masses up to about that of glucose (180 Da), which is still quite small compared with the molecular mass of a protein or a DNA molecule, which may range from several thousand to several hundred thousand daltons. Haemoglobin, the oxygen-carrying protein in red blood cells has a relative molecular mass of about 258,000 Da. One of the components of starch, the multiple-branched amylopectin, is believed to be the largest naturally occurring molecule and, depending on the plant source, may have a relative molecular mass up to 10 million Da.

All biopolymers have a property called directionality, meaning that they are not the same when 'read' from different ends. This property will become obvious in the following discussion. The main concern of this chapter is the proteins, and especially the enzymes.

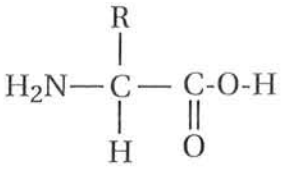
The word protein is derived from the Greek *proteios*, meaning 'primary', as it has long been considered that these large molecules were essential components of all life. They were almost certainly not, however, the first potentially biological macromolecules to be formed on Earth. This role probably fell to the nucleic acids, most likely RNA, but this will be dealt with later. I use the word 'potentially' as the formation of many kinds of molecules preceded the emergence of life. Some of these precursor molecules were the relatively simple, stable monomers that eventually became joined to form the first polymers.

Proteins have a number of roles in living organisms, including structural (collagen in connective tissue and tendons, keratin in hair and feathers), transport (haemoglobin carries oxygen, some lipoproteins carry fats and cholesterol), messenger (certain peptide hormones act as molecular signals), catalytic (enzymes), regulatory (insulin controls entry of glucose into cells), protective

(immunoglobulins of the immune system), contractile (actin and myosin of muscle), and storage (albumin in egg, casein in milk).

Our major focus in this chapter will be the enzymes, which illustrate many of the amazing variety of properties which proteins may exhibit, and which are intimately involved in metabolism and its control.

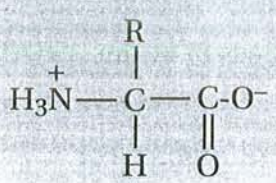
The monomers that are strung together to form proteins are the amino acids. There are about 20 amino acids that appear commonly in proteins, plus a few others that are not so widely used in nature. It is possible to draw a common structural formula for the type of amino acids found in proteins:



The H_2N- is an α-amino group (it is on the carbon α to the COOH) and the COOH is an acid (carboxyl) group, hence the name α-amino acid. The R in this case stands for any one of the 20 side-chains that make each amino acid different from the rest. I won't give the structures for all of these here, but refer you to Appendix E.

Box 8.1 Some properties of α-amino acids

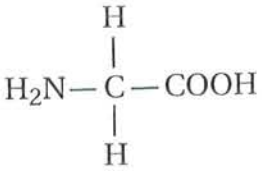
At physiological pH values, the amino acids occur not in the form illustrated above, but as dipolar ions, called zwitterions:



The zwitterion arrangement occurs because the pK values of carboxyl groups occur in a small range around 2.2, and the pK values of the amino group occur around 9.4. This means that at pH ~7, both groups are more than two pH units away from their pKs, and will be effectively fully ionized. The pK values given in Appendix E refer to the pK_1 (for COOH groups) and pK_2 (for NH_2 groups). Other ionizable side-chains also have their pK values listed. Amino acids can act as both acids and bases. In free form, or in peptides and proteins, they are important in maintaining the pH in their particular cellular environment. The actual pK of an ionizable group will depend on its surroundings. In a protein, for example, amino acid side-chains that lie far apart in the primary sequence may be close neighbours in space due to folding of the protein backbone. Thus $-COO^-$ and $-H_3N^+$ in close proximity can have their pK values moved by one pH unit or more. Other close neighbour groups can also exert an influence. The ability of the side-chains of amino acids to gain or lose charge, for example by proton exchange, explains their presence in the active sites of enzymes, where they often act as acid/base catalysts (see Figure 8.15).

I'll mention certain ones where necessary. The side-chains vary in their chemical and physical properties, from acidic through neutral to basic, and from hydrophilic to hydrophobic. This range of side-chain properties plus the fact that they can be arranged in any order, provides for a vast array of protein properties.

The simplest side-chain is a hydrogen atom, so the simplest amino acid, glycine (abbreviation: Gly) has the structure



Two amino acids may be covalently linked by joining the amino end of one to the carboxyl end of the other. For two glycines, it is possible to 'eliminate' a water molecule, in a typical condensation reaction, to yield the two glycines linked as follows:

$$H_2N-CH_2-C\mathbf{OOH}+\mathbf{H}_2N-CH_2-COOH \longrightarrow$$
$$H_2N-CH_2-\mathbf{CONH}-CH_2-COOH+\mathbf{H_2O}$$

The **−CONH−** part of the new molecule links the two glycines together by a peptide linkage, one of the most important bonding types in biochemistry. The new molecule is called a dipeptide as it consists of two amino acids. Its specific name is glycylglycine, reflecting its amino acid composition. In similar fashion we may have tripeptides, tetrapeptides, and so on, up to hundreds of amino acids. Above 10 or so amino acids, the term polypeptide is used. In nature, peptide bonds are formed by enzymes, on the cellular structures called ribosomes, as outlined in Chapter 9. Peptide bond formation is endothermic, the energy being provided by ATP via various coupling reactions. The chemical and physical nature of the side-chains represented by R varies widely, leading to a variety of properties when the amino acids are joined to make a protein chain. This ensures that a plethora of useful properties is available in proteins. The amino acids may be joined in any gene-coded sequence, so with 20 possibilities to choose from at any position in a protein chain, there is an enormous number of unique sequences available for the formation of proteins. As an illustration let's consider joining any two of the 20 amino acids to form a dipeptide.

The first position of the dipeptide can be taken by any of the 20 amino acids, and so can the second position. The number of possible dipeptides is thus $20^2 = 400$. There are 20^3 or $20 \times 20 \times 20 = 8000$ possible tripeptides, and so on. The number of possible peptides soon becomes mind boggling. There is almost an unlimited number of protein chains possible, each with a unique amino acid sequence. For example, there are in theory, 20^{50} different peptide chains 50 amino acids long, but nowhere near this many types actually exist.

It is the sequence of amino acids in the chain, plus the number of amino acids, which *ultimately* determine shape and the properties of an individual protein. There is more to this statement than meets the eye because functional proteins

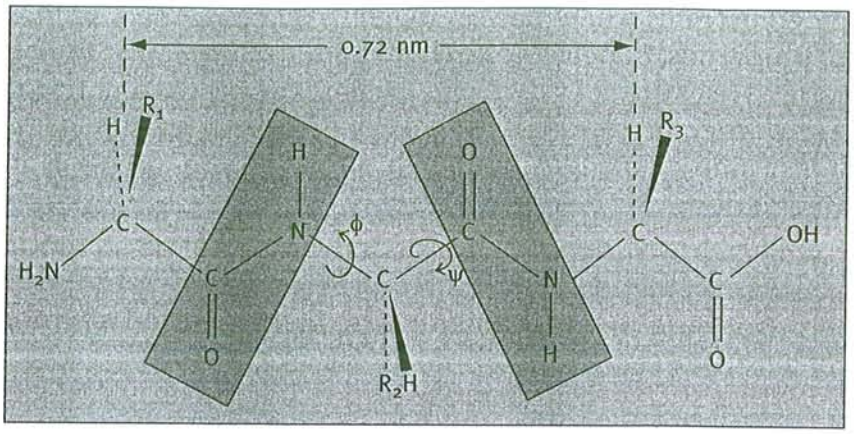


Figure 8.1. A tripeptide formed from amino acid residues having side-chains R_1, R_2, and R_3. Note the rigid, planar peptide linkages (shading) and the bonds ϕ and ψ that allow relatively free rotation. The fully extended β-conformation is shown.

are not merely long chains of amino acids strung together like beads on a string—far from it. They can adopt a large variety of three-dimensional shapes, or conformations. Figure 8.1 shows the bonds that confer flexibility on peptide and protein chains. Depending on the side-chains involved, there is relatively free rotation allowed around the bonds marked ϕ and ψ. There is a ϕ and ψ angle for each amino acid in a protein chain.

There are four well-recognized levels of complexity in proteins: the primary, secondary, tertiary, and quaternary structures (Figure 8.2).

The primary structure of proteins is the amino acid sequence, for example Gly–Val–Asp–Glu–Tyr–Cys–Ser.

Secondary structure refers to the folding of the primary chain into a number of shapes. There are two major regular shapes, or ordered conformations. The best-known ordered conformations in protein secondary structure are the alpha helix (α-helix) and the beta-strand (β-strand) (Figure 8.2(b)). The latter is also termed the β-sheet or β-pleated sheet.

Tertiary structure involves further folding of the protein chain, which when completed may have some parts of its length in the α-helix and/or some in the β-strand form, all folded into a three-dimensional shape (Figure 8.2(c)). The muscle protein myoglobin in its native form is folded into a compact globular shape, and this fact is reflected in the name (Figure 8.3). Myoglobin is very similar in overall shape to the individual chains of haemoglobin.

Myoglobin contains eight α-helical regions (A to H in Figure 8.3) separated by bends or kinks in the polypeptide chain (AB, BC, etc.) but no β-pleated sheets. About 75% of the chain is α-helical. Myoglobin is extremely compact, there being very little 'free space' inside, much of which is occupied by non-polar side-chains, in accordance with the hydrophobic effect.

Proteins may also adopt a quite linear (elongated) overall ordered conformation, such as that found in the fibrous silk protein, fibroin, which consists largely of β-sheets (Figure 8.2(b)). The arrangement of protein chains in fibroin is stacks of antiparallel β-sheets. The amino acid side-chains (−R) protrude alternately above and below the plane of the sheet. In fibroin, there are many

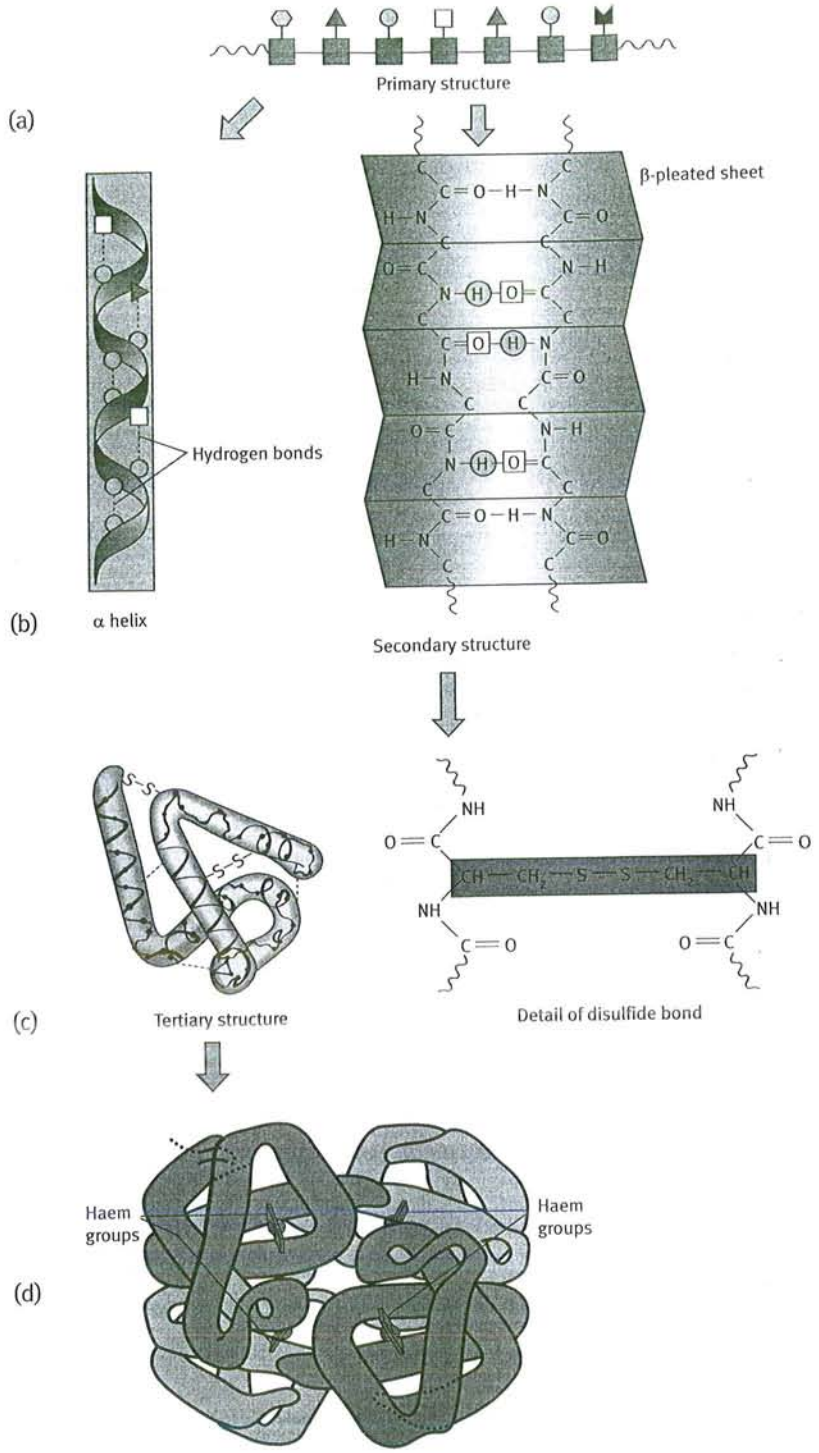


Figure 8.2. The four levels of protein structure, illustrated by stages in the formation of haemoglobin. (a) Primary structure is the amino acid sequence. (b) Secondary structure involves the ordered conformations of the α-helix and the β-pleated sheet. Secondary structure is stabilized by hydrogen bonds and other secondary bond types. (c) Further folding of secondary structures leads to the tertiary structure, which is often stabilized by disulphide bonds. One of the four subunits of haemoglobin is shown. (d) Haemoglobin formation is completed by the association of four subunits to form the active quaternary structure. (Illustrations (a) to (c) from Talaro, K. and Talaro, A. *Foundations in Microbiology*, 2nd edn, p. 47. (c) 1996 Times Mirror Higher Education group, Inc. Illustration (d) from Irving Geis. Image from Irving Geis Collection, Howard Hughes Medical Institute. Rights owned by HHMI.)

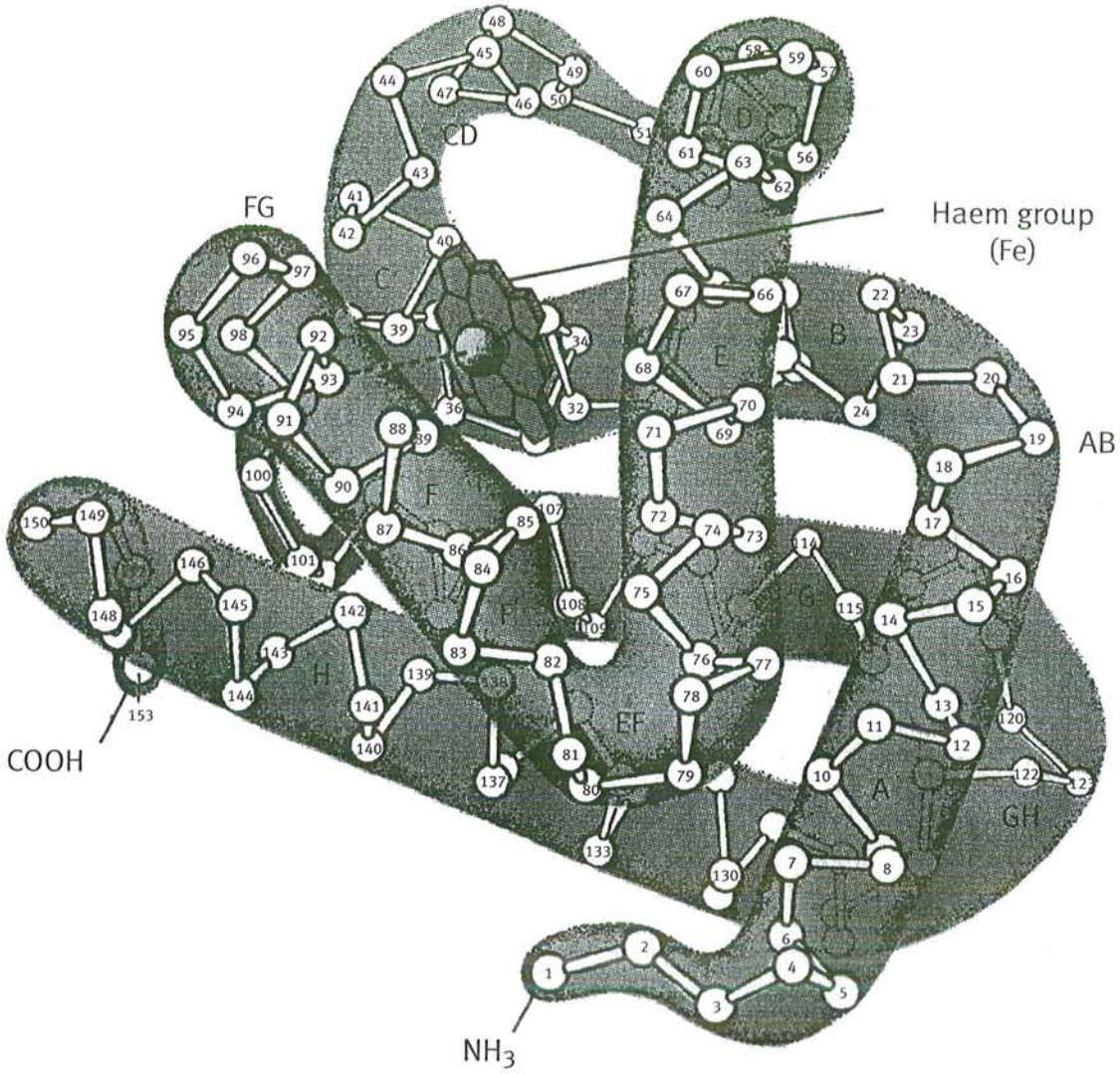


Figure 8.3. Structure of myoglobin, showing the secondary and tertiary structure (Illustration from Irving Geis. Image from Irving Geis Collection, Howard Hughes Medical Institute. Rights owned by HHMI.)

regions where glycine (R = H) alternates with either alanine ($R = CH_3$) or serine ($R = CH_2OH$). The glycines are arranged on one side of the sheet and the alanines/serines on the other, allowing two sheets to pack closely together—glycine to glycine, or alanine/serine to alanine/serine. About 80% of fibroin consists of glycine, alanine, and serine (Figure 8.4).

The keratins of bird feathers are also made up of stacked β-sheets. Such a regular, tightly packed arrangement allows stabilizing secondary bond forces to operate maximally. The result in silk is a fibre with high tensile strength and resilience.

Another fibrous protein, collagen, found in cartilage and connective tissue, consists of three α-helical chains, each formed into a superhelix and intertwined (Figure 8.5). Collagen has an unusual amino acid composition, made up of multiple repetitions of the sequence glycine, proline, and hydroxyproline.

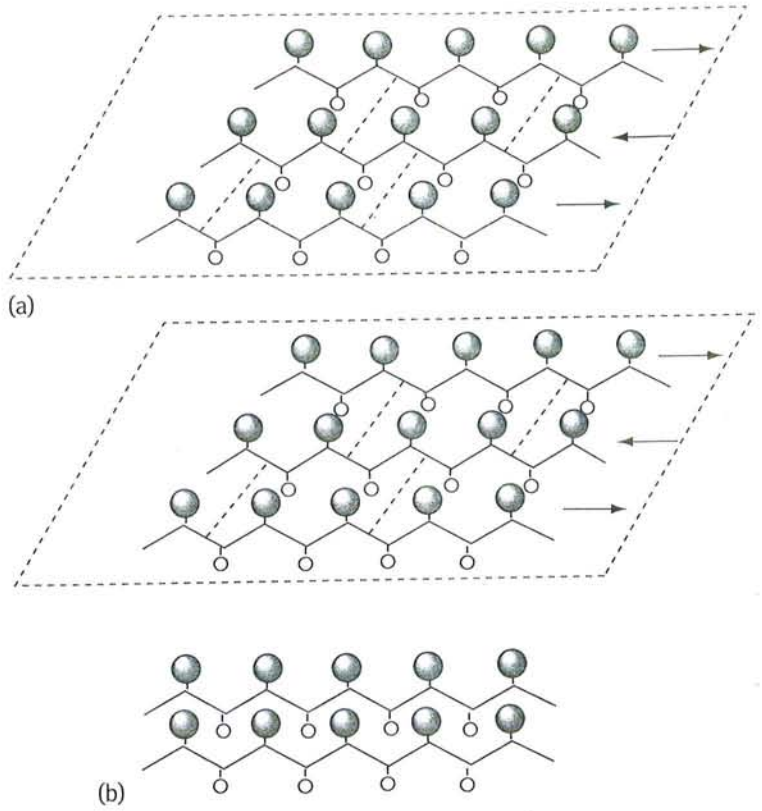


Figure 8.4. The structure of silk fibroin. (a) Two layers of fibroin showing antiparallel β-sheet arrangement of the chains. The vertical distance between sheets 1 and 2 is greatly exaggerated for clarity. Interchain hydrogen bonds between −C=O and −HN− groups in the protein backbone are also shown. (b) Front-on view of layers 1 and 2 showing the intercalation of large sidechains ● (alanine and serine.) with the H sidechains of glycine (○).

Hydrogen bonds occur between −OH groups on side-chains of amino acids in the helices. Low hydroxylation can lead to scurvy. Adequate hydroxylation levels in mammals require an adequate supply of vitamin C in the diet. Covalent bonds (not shown) help to hold the three superhelices together.

Other proteins consist of several individual chains, or subunits, held together by secondary forces to form quaternary structures. The haemoglobin (Hb) molecule is an example. Each protein has an overall structure suited to its biological function. As proteins come off the ribosome in a linear fashion, they must be somehow folded into this active shape or conformation. Figure 8.2 illustrates this process schematically for haemoglobin. Protein folding and its energetics are discussed later in the chapter.

At each of the structural levels described above, maintenance of that structure is important for the function carried out by the protein.

Some obvious questions arise:

1) What determines the three-dimensional shapes a given protein chain may adopt?
2) What causes it to adopt such a shape?
3) What maintains the shape once it has been achieved?

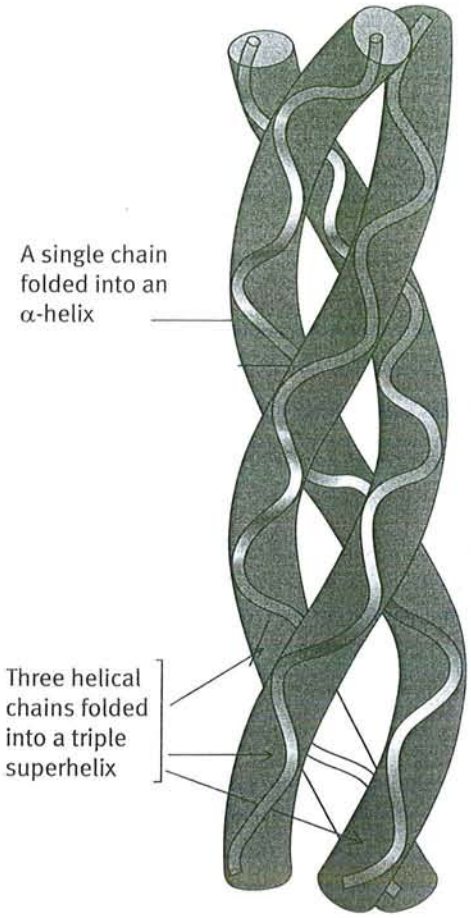


Figure 8.5. Collagen: a triple superhelical protein found in skin, bone, tendon, cartilage, and eye lens. (Illustration from Irving Geis. Image from Irving Geis Collection, Howard Hughes Medical Institute. Rights owned by HHMI.)

The answer to each of these questions involves energy and includes a number of principles already discussed. The discussion below is restricted mainly to the proteins that form enzymes or enzyme systems.

1) The amino acid sequence (primary structure) of a protein is the major determinant of the shapes it may adopt by folding of the protein chain. Also involved will be the precise environment within the cell, for example whether the protein will be free in solution, inserted into a lipid membrane, or be part of a structure such as a ribosome or a gene.
2) What causes the protein chain to adopt a particular shape? How is energy involved to bring the folding into being, and to maintain it?

Consider the synthesis of the protein chain that will become a globular myoglobin molecule. It comes off the ribosome in a linear fashion, like a string of beads, and (eventually) adopts its secondary structure with a high proportion of α-helical regions. The α-helical regions are joined by folds and bends, rather like a long sausage folding into a heap, and finally form into the compact globular tertiary structure. The millions of myoglobin molecules in

mammalian muscle all fold unerringly into the same shape. The process must therefore be thermodynamically favourable. A net driving energy (negative Gibbs energy, ΔG) is involved overall, even though there is a decrease in entropy as the folding of the chain makes it more ordered. What are the origins of this negative ΔG?

For the overall process of folding

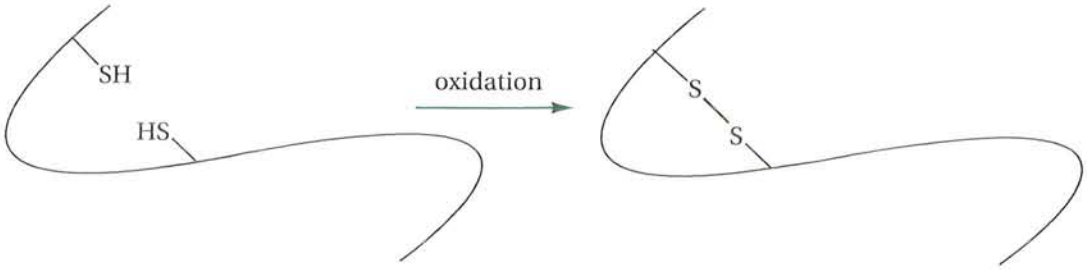
$$\Delta G = \Delta H - T\Delta S$$

A contribution to ΔG that favours folding comes from the formation of hydrogen bonds between adjacent parts of the polypeptide chain during the formation of α-helices, which are apparently formed early in the folding process (see the α-helix in Figure 8.2(b)). Some favourable energy derives from the association between oppositely charged $-$R groups, such as $-COO^-$ and $-NH_3^+$, or coordination of metal ions such as Mg^{2+} or Ca^{2+} to negatively charged side-chains.

Calculations have shown that for the polar side-chains, the entropy and enthalpy terms approximately cancel out, so that $\Delta G_{\text{folding}} = \sim 0$. For the non-polar side-chain contribution, the ΔH and $T\Delta S$ terms are positive, and are thus unfavourable for folding. The solvent water now plays an important part in folding energetics. Large numbers of water molecules around non-polar groups in the unfolded state are restricted in motion (see Chapter 7). When these are 'liberated' as the hydrophobic groups fold inside, there is a large increase in entropy, which compensates for the repulsive energy and makes the overall folding process favourable. The largest contribution to the stability of a folded protein derives from this entropy change. Interestingly, the ΔG for folding is typically not large, in the -20 to -40 kJ mol^{-1} range for the macromolecule, leading to the conclusion that a folded protein is barely stable, that is its native conformation is somewhat flexible and is often easily denatured. The flexibility has important implications for protein dynamics, especially in the case of enzymes, as discussed below. The propensity for denaturation needs to be tolerated, and could be considered as the price paid for the subtleties of protein regulation. Even natural selection can't win all the time. (In hyperthermophiles, natural selection has, once more, done very well. The amino acid sequences 'chosen' for their enzymes are such that high temperatures (100°C plus) can be tolerated.)

A fully folded, water-soluble globular protein usually has the majority of its polar side-chains facing out towards the aqueous environment, with the hydrophobic side-chains buried deep inside, well shielded from being able to disrupt the structure of the surrounding water. Despite this general principle, there are many examples of water molecules also being found inside globular proteins. They may have some specific role in forming bridging hydrogen bonds between polar groups or be involved in proton-catalyzed reactions. The globular structure minimizes the volume in contact with the water environment, thus maximizing the number of energetically favourable hydrogen bonds in the surrounding water. Disulphide bonds between two cysteine residues in different parts of the protein chain may also be involved. These are covalent bonds that, when in place, restrict the folding capability of the protein chain, so are therefore important in determining, stabilizing, and maintaining its overall three-dimensional shape. The $-$R group in the

amino acid cysteine is $-CH_2SH$. Schematically, the formation of a disulphide bond between two cysteine groups in a protein chain can be represented as:



In vivo, the oxidation of the two $-SH$ side-chains is usually carried out by a specific enzyme. A recent review points out the importance of disulphide bond formation in mitochondria and the endoplasmic reticulum system (Riemer *et al.* 2009). The above description is a simplified account of protein folding. In principle, there are many tertiary conformations that a linear molecule can adopt. The only requirement is that the ΔG for the entire folding process be negative. There can be many conformations of roughly comparable energy level that satisfy this criterion. This leaves plenty of scope for some proteins to adopt a non-biological conformation. Inside cells, protein folding processes can be complex. As the correct folding of a protein chain is so important, considerable research effort has been put into studying this process, and three protein-folding mechanisms have been identified (Figure 8.6).

Some proteins fold without assistance—they spontaneously adopt the correct conformation on thermodynamic grounds alone (Figure 8.6(a)). Others may fold

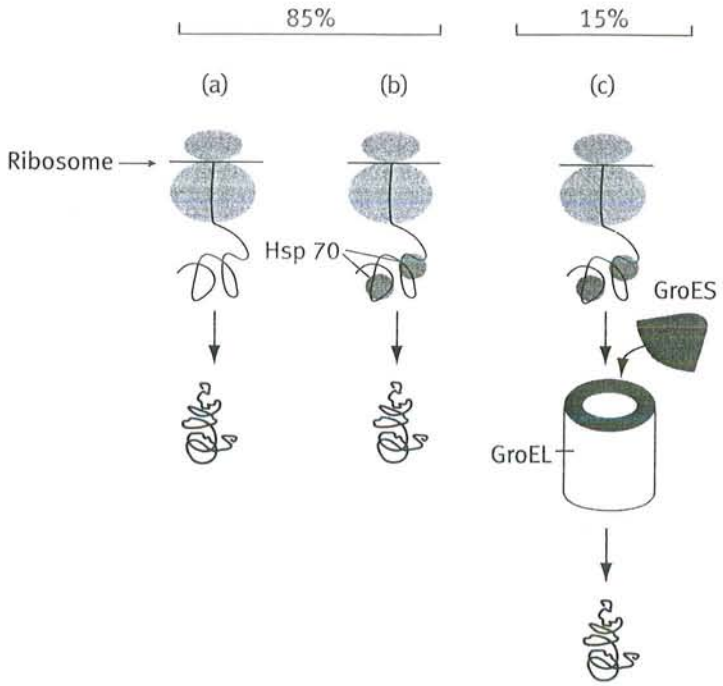


Figure 8.6. Mechanisms of protein folding. (Adapted from Netzer, W.J. and Hartl, F.U. (1998) Protein folding in the cytosol: Chaperonin-dependent and -independent mechanisms. *Trends in Biochemical Sciences* **23**, 68–73.)

incorrectly unless they interact with a chaperone, such as Hsp70 (heat-shock protein 70) (Figure 8.6(b)). The process is driven by ATP hydrolysis. About 85% of proteins fold as shown in Figure 8.6(a) and (b). The rest are folded with the aid of other proteins called chaperonins (Figure 8.6(c)). Chaperonins, large structures consisting of protein subunits, form a cage around each newly formed protein chain. Protein folding, which takes place inside the cage, is ATP dependent. Up to 100 ATP molecules must be hydrolyzed to complete protein folding. This is a large expenditure of energy. For the cell to invest such an amount of energy in this process (plus the associated genes, their control, and synthesis of the subunit proteins) it must be very important. As we have seen, natural selection is parsimonious and tends to waste nothing. In fact, there is good evidence that if a correctly functioning, properly folded protein somehow loses its native conformation, the result for the organism can be disastrous. Prions (proteinaceous infectious particles) are proposed to be particles consisting of proteins that have adopted an abnormal conformation. They appear to be a prime cause of some diseases of the central nervous system. In humans, Creutzfeldt–Jacob disease and kuru are fatal degenerative diseases. In other mammals the diseases scrapie in sheep and BSE/mad cow disease, have similar origins. Although apparently inherited, according to some evidence, no nucleic acid involvement has been shown. Prions may be acquired by infection, perhaps by ingesting food from an infected animal. The prions are believed to differ from the natural proteins only in their secondary and tertiary structures. One proposal is that the natural form (PrP^{c}) has a large proportion of α-helices, whereas the the scrapie form (PrP^{sc}) consists of both α-helices and β-strands. Apparently, the presence of PrP^{sc} can induce the normal PrP^{c} to change to the pathogenic conformation, forming plaques of ever-increasing size that eventually develop into the disease. The plaques ultimately destroy tissue in the central nervous system.

In the normal course of protein maturation, the folded, active protein structure is maintained by the bonds and forces that stabilize the tertiary and quaternary structure (Figure 8.7).

Hydrogen bonding was described in some detail in Chapters 5 and 7 as an example of secondary bonds, the weak interactions that are important in biomolecules. Hydrophobic interactions were also covered in Chapter 7. Secondary bond strengths relative to some covalent bonds are quite low, but their influence can be large, as many secondary bonds can act cooperatively:

Type	Strength (kJ mol^{-1})	Comments on strength
Covalent bonds	Range 200–900	Strong primary bonds
Hydrogen bonds	0–30	Varies with bond polarity; directional
Van der Waals	0.4–4.0	Close contact/large area enhances strength
Ionic bonds	20	Depends on polarity of the + and − species
Hydrophobic interactions	4–12	Highly dependent on environment, especially the effect on water structure

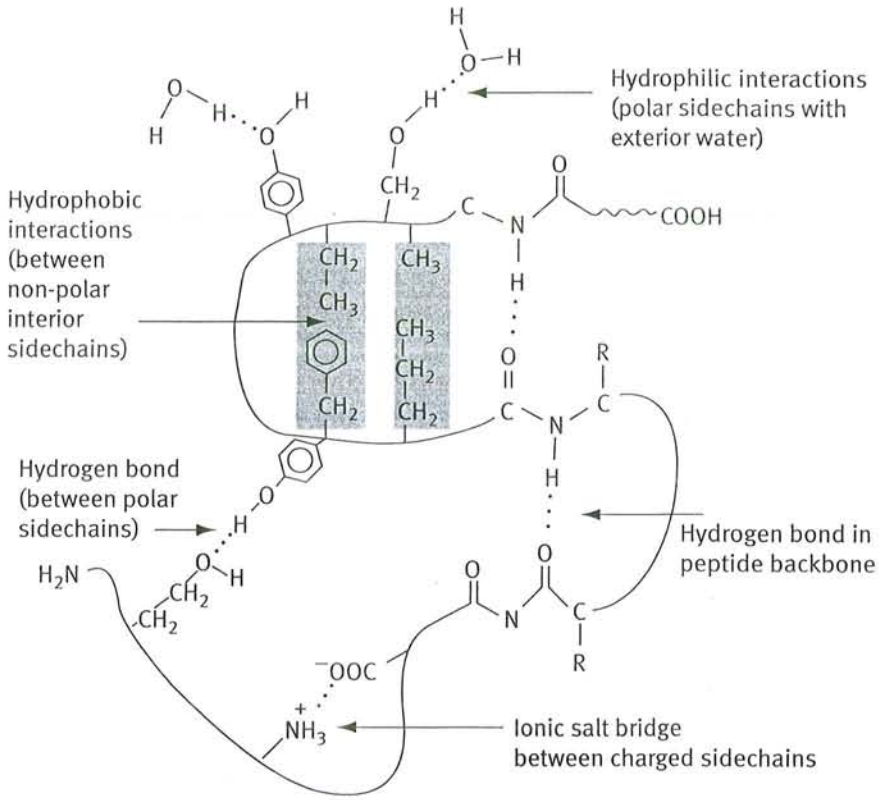


Figure 8.7. Secondary forces that stabilize protein structure. (Illustration adapted from Crowe, J., Bradshaw, T., and Monk, P. *Chemistry for the Biosciences*, p. 106. © 2006 Oxford University Press.)

Van der Waals interactions occur between all molecules, and are very short range. They are the result of induced electrical dipoles involving electrons of molecules in close contact. Their strength is enhanced by complementarity of shape and area of contact between the molecules involved. These conditions can occur when substrates bind to enzymes. An example is the binding of the enzyme lysozyme to a carbohydrate substrate, where the attractive van der Waals energy has been calculated to be about 60 kJ mol^{-1}, a considerable value (Garrett and Grisham 1999).

Many proteins, especially enzymes, undergo slight but functionally important conformational changes when carrying out their biological roles. They can change shape. Haemoglobin is one of the most fascinating examples of this. Each of the four subunits of haemoglobin has a globular shape similar to that of a single myoglobin molecule. Both molecules possess a haem prosthetic group, a structure containing a ferrous iron (Fe^{2+}) ion that binds molecular oxygen, O_2. The four haem groups in haemoglobin allow it to act as an oxygen carrier in red blood cells. As venous blood, low in oxygen and high in carbon dioxide collected from working cells, circulates in the alveoli of the lungs, carbon dioxide is lost from the red cells and is exhaled. As each new breath of air is taken, the partial pressure of oxygen in the lungs becomes quite high and oxygen is taken up by the haemoglobin of red cells passing through capillaries very close by. The blood continues to the heart, to be transported around the body. As the oxygenated blood passes 'hard-working' cells where metabolic activity has depleted the oxygen during the processes of

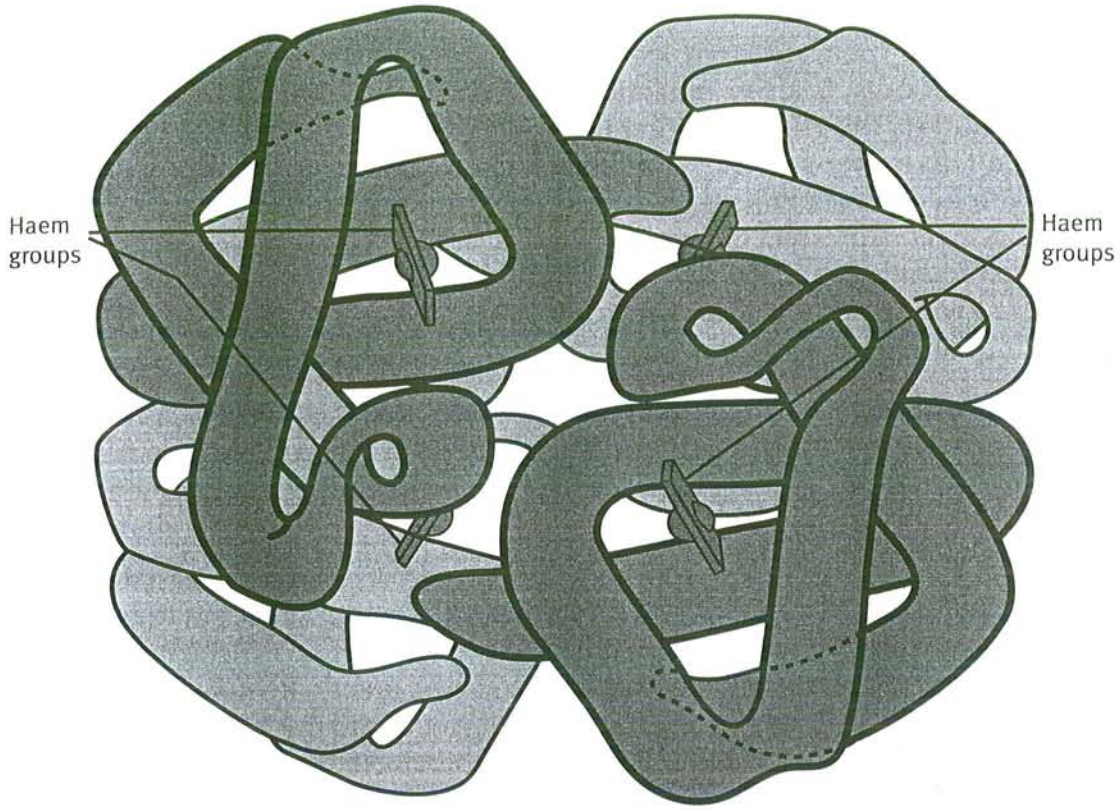


Figure 8.8. Quaternary structure of haemoglobin. (Illustration from Irving Geis. Image from Irving Geis Collection, Howard Hughes Medical Institute. Rights owned by HHMI.)

respiration, the haemoglobin releases most of its oxygen to the cells and picks up carbon dioxide in the form of bicarbonate (HCO_3^-) (Figure 8.8).

The relative positions of the four chains alter when oxygen binds, indeed crystals of haemoglobin shatter when exposed to oxygen, a dramatic demonstration of the effect. Studies by X-ray crystallography have revealed that oxyhaemoglobin and deoxyhaemoglobin differ significantly in their quaternary structures. The precise movements and their distances are known in detail. Why should haemoglobin behave in this way? The biological significance is great. The role of haemoglobin in mammals is to efficiently bind oxygen from the lungs, transport it to the tissues, then efficiently release it. I emphasize the binding and release processes, as these appear to be contradictory properties in the same molecule. If oxygen is efficiently bound by haemoglobin, how can it also be released efficiently later? The answer is by making use of the allosteric effect. Put concisely, the movement of Hb subunits when the first oxygen molecule binds, assists the binding of the second and subsequent oxygen molecules (haemoglobin can bind four oxygen molecules, one per subunit). This process is a form of cooperative binding, and is typical of allosteric effects. Allosteric means 'in another place' ('other space' in Greek). In haemoglobin and in the case of allosteric enzymes, the term allosteric indicates that the effects are caused by molecules or ions binding away from the active site.

When the fully oxygenated haemoglobin (HbO_2) reaches working tissues, the environment is quite different from that in the lungs. The pH, particularly in active muscle cells, is lower as lactic acid is produced. Protons have an affinity for, and bind to, the HbO_2, altering the spatial relationship between the α- and β-chains. This movement causes a shape change that lowers the affinity of HbO_2 for oxygen and assists in its release, leaving protons bound to Hb.

Also, as an end-product of aerobic metabolism, carbon dioxide will be present in high concentration in the active tissue. Because the pH of blood is about 7.4, most CO_2 will be in the form of bicarbonate, HCO_3^-, because at pH 7.4 the reaction below is driven towards the right:

$$CO_2 + H_2O \rightleftarrows H_2CO_3 \rightleftarrows HCO_3^- + H^+$$

Protons formed in this way further assist the release of oxygen from HbO_2. The HCO_3^- is transported in the blood back to the lungs. When our oxygen-depleted Hb arrives back in the oxygen-filled lungs, its bound protons are released and Hb combines with fresh oxygen. In the reverse of the reaction shown in the equation above, the released H^+ combines with HCO_3^- to form H_2CO_3, which dissociates into water plus CO_2, both of which are exhaled in the breath. Removal of the CO_2 in this manner helps to drive the reversible reaction above to the left. A small amount of CO_2 binds to Hb directly and is transported thereby to the lungs and released.

To complete the picture of this remarkable molecule, another metabolite bisphosphoglycerate (BPG, see Chapter 11) also has a marked effect on the oxygen-carrying capabilities of Hb. When oxygen-laden Hb reaches the cells, it needs to release the four bound O_2 molecules efficiently. The highly charged BPG is present in about a 1:1 ratio with Hb, and binds strongly to its allosteric site, in the interior between the two β-chains, stabilizing the deoxy form and promoting the release of O_2. BPG plays an important role in supplying oxygen to a growing mammalian foetus. Foetal Hb (HbF) binds BPG less efficiently than adult Hb, with the result that HbF has a greater affinity for oxygen than the mother's Hb. This ensures that oxygen transfer from mother to foetus is favoured. Why should foetal Hb bind BPG less strongly? In maternal Hb, the negatively charged BPG binds to the positively charged histidine side-chains at position 143 in the β-chains. Foetal Hb possesses a different equivalent to the adult β-chains, called γ-chains. These γ-chains have serine at position 143 instead of the basic histidine. The serine side-chain is $-CH_2OH$, is neutral, and does not become positively charged, hence removing two positive charges for binding to BPG. The result is that foetal Hb still binds oxygen at low partial pressures of O_2, whereas adult Hb will have lost it.

The behaviour of Hb is known in great detail, much of which has not been dealt with above. All I wish to add is that the precise mechanism of oxygen uptake and binding, plus the movements of the four subunits, are known down to the exact amino acid side-chains involved, including the location of those involved in important hydrogen bonds, the location of salt bridges (some of which involve chloride ions Cl^-), and the location of others which form between $-H_3N^+$ and $-COO^-$ groups of amino acids in the chains. The disruption of some of these secondary bonds during oxygen uptake is energetically favourable

overall. Although individually small in magnitude, they are collectively and essentially involved in the binding and subsequent release of oxygen. The energy contributions of all these secondary bonds are vital to the function of haemoglobin, and are thus vital to our own survival. The quaternary structure of Hb, with its four subunits, has evolved to allow the emergence of a property, cooperativity, of which a single polypeptide chain is incapable—another example of the art of the possible. The allosteric effect is a major mechanism by which enzyme activity is regulated (see below).

I have treated the principles of protein three-dimensional structure in some detail because of its close relationship to function. In general, if a protein is 'denatured', that is if the native conformation is somehow destroyed, its natural function is also likely to be destroyed. This is particularly the case with enzymes. Most of the forces holding proteins in their native conformations are relatively weak secondary forces, and as such are individually easy to disrupt. If enough of its hundreds of hydrogen bonds, electrostatic forces, and van der Waals forces are disrupted, a typical protein will be forced to change its tertiary structure, often drastically. Familiar examples include the cooking of an egg or a piece of steak: enzymes become useless, the white and yolk solidify, collagen in the steak shrinks and loses water. The changes in properties of these high-protein foods brought about by cooking is obvious to the naked eye. It is possible to demonstrate that under carefully controlled conditions denaturation is not necessarily irreversible. During experiments to demonstrate that the native shape of a protein is determined by its primary structure, Christian Anfinsen was the first to demonstrate that the refolding of a denatured protein (the enzyme ribonuclease) in vitro could be a spontaneous process. Put briefly, after denaturation and allowing some time for renaturation under appropriate conditions, the ribonuclease was found to have its enzymic activity restored. When in the denatured form, it was completely inactive.

How do proteins become denatured? Heat is the most familiar cause. Heat energy, for example that involved in cooking, is sufficient to disrupt the secondary bonds in most proteins. Changes in pH and ionic strength, plus treatment with detergents, may also cause denaturation by changing the patterns of hydrogen bonds and ionic bonds. Some proteins in solution are denatured by vigorous mixing. Denatured proteins may collect at the air/water interface as froth or foam. Egg white when beaten forms a protein foam that is stabilized by denatured ovalbumin. Cooking causes further denaturation, which hardens the surface of the foam and the result (for a skilful cook) is a light shell used for the delicious dessert known in Australia as Pavlova. Further examples of deliberate denaturation of proteins for purposes of customer appeal include beer pouring and cappuccino coffee making. Blasted into existence by a jet of steam, cappuccino froth consists largely of heat-denatured milk proteins. Some addicts almost swoon with delight at the structures and odours so generated. Beer drinkers can become passionate about the 'head' of their favourite brew, although it is merely a foam stabilized by a little denatured protein (and perhaps an additive) produced when pressurized carbon dioxide is suddenly released. Try telling that to a crowded bar full of Guinness drinkers on a Friday night!

8.2 Some very special molecules: enzymes

As we have seen in Chapter 5, a catalyst is a substance that increases the rate of a chemical reaction by altering the reaction pathway, thus lowering the overall activation energy. In biological systems, some reactions would be impossibly slow, and cause a log-jam in metabolic pathways, without the presence of enzymes to speed them up. In some inherited metabolic diseases, a crucial enzyme is missing (or present in low amounts), leading to just such log-jams. An example is the genetic disease phenylketonuria, which I will discuss later in the chapter. Like all catalysts, enzymes change the rate of a reaction without being changed permanently themselves. This latter property is important, as a single enzyme molecule may be used many times over. Enzymes are usually highly specific for the reaction they catalyze, and their activity can be regulated. Essentially, all enzymes are proteins. Some enzymes are able to carry out their catalytic functions using their protein structure alone. Others need the help of cofactors. Cofactors can be metal ions or organic molecules called coenzymes, such as the nicotinamide adenine dinucleotide (NAD) or nicotinamide adenine dinucleotide phosphate (NADP) required in electron transfer reactions. We know many of the precursors of coenzymes as vitamins, about which more will be said later. Other enzymes may need the presence of a specific metal ion, such as Ca^{2+}, Mg^{2+}, or Zn^{2+}, to be active. A cofactor that is covalently linked to the protein is called a prosthetic group. Although not enzymes in the accepted sense of the word, some catalytically active RNA molecules have been identified. These might, and I stress *might*, have had a role in the early development of life (Luisi 2006).

Enzymes certainly increase reaction rates, but they do much more as well. A particular enzyme is in general highly specific for its particular substrate and usually will not catalyze reactions with other substrates. For biochemical reactions, embedded in biochemical pathways which need to be strictly controlled, this is a matter of necessity.

A suite of enzymes exerts control over a coordinated series of reactions making up a particular pathway, allowing the pathway to adapt to changing conditions. If there is a demand for more ATP, as during heavy exercise, the pathways involved in respiration, glycolysis, the TCA cycle, and the electron transport chain (ETC) of oxidative phosphorylation are brought into high activity. When demand falls, feedback mechanisms, mediated by key enzymes in the pathways, reduce the overall activity back to normal levels.

There are some important exceptions to the specificity rule, such as a number of enzymes found in mammalian liver. The liver is the site of detoxification of foreign substances potentially harmful to the body. Enzymes that catalyze oxidations, reductions, and desulphurations are examples of the first phase of detoxification reactions in the liver. Many of the oxidations are performed by cytochrome P450 systems, which have a broad substrate specificity. Obviously the liver doesn't 'know' what unwanted compounds are likely to be ingested by the mammal, so the ability to cope with a range of possibilities is a distinct advantage.

How do enzymes perform all the functions we have mentioned? Let's take them one at a time, starting with the catalytic properties. How is it that a protein,

present in low concentrations, can increase the rate of a reaction by a factor of a staggering 10^{12} or more? Such increases in reaction rate are far more than man-made, synthetic catalysts can achieve. Let's take the familiar biological oxidation of glucose as an example:

$$C_6H_{12}O_6 + 6CO_2 \longrightarrow 6CO_2 + 6H_2O \quad \Delta G^{0\prime} = -2871 \text{ kJ mol}^{-1}$$

This is a large amount of energy, but we are considering thermodynamics alone. The thermodynamics tells us nothing about the rate at which the reaction will occur. One could leave glucose (or just about any carbohydrate) lying around at room temperature, in contact with the oxygen in the air, for months and it would remain essentially unchanged. That situation is obviously unacceptable for living cells, which may need the energy immediately. The series of enzymes involved in glucose metabolism allows the energy to be released more rapidly in a controlled way that allows the regeneration of ATP, the major immediate source of cellular energy. Enzymes give the cell kinetic control over the thermodynamic potential of glucose.

What actually happens when an enzyme becomes involved in a reaction such as (8.1)?

$$W + X \longrightarrow Y + Z \tag{8.1}$$

In an uncatalysed biochemical reaction the reactants, molecules W and X, are surrounded by a hydration shell of water molecules, and they move about in solution in a random fashion. Only rarely will W and X collide with sufficient energy, and in the correct orientation, for a collision complex (X–W) to occur (Figure 8.9).

Before the products, molecules Y and Z, can form, the collision complex must pass through a transition state, which can be thought of as orienting W and X in such a way and with sufficient energy to facilitate any bond making/bond making necessary to form Y and Z. It has been suggested that one important aspect of enzyme catalysis is the ability of the reactive site to organize the substrate into a near-attack conformation.

The near-attack conformation puts the atoms that will react in close contact, in the correct orientation for bond breaking and making. It has been estimated that near-attack conformations leading to the transition state occur about 0.0001% of the time in uncatalysed reactions (Bruice 2002). Formation of the transition state requires a relatively large activation energy (ΔG^{*}). In the uncatalysed reaction, very

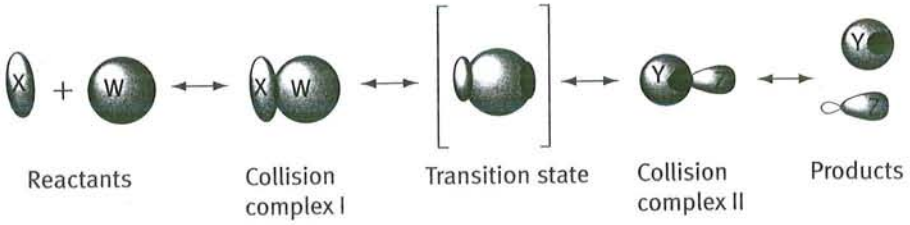


Figure 8.9. Mechanism of an uncatalysed reaction. Collisions leading to a viable transition state are rare. Near-attack conformations leading to the transition state occur about 0.0001% of the time (Bruice 2002).

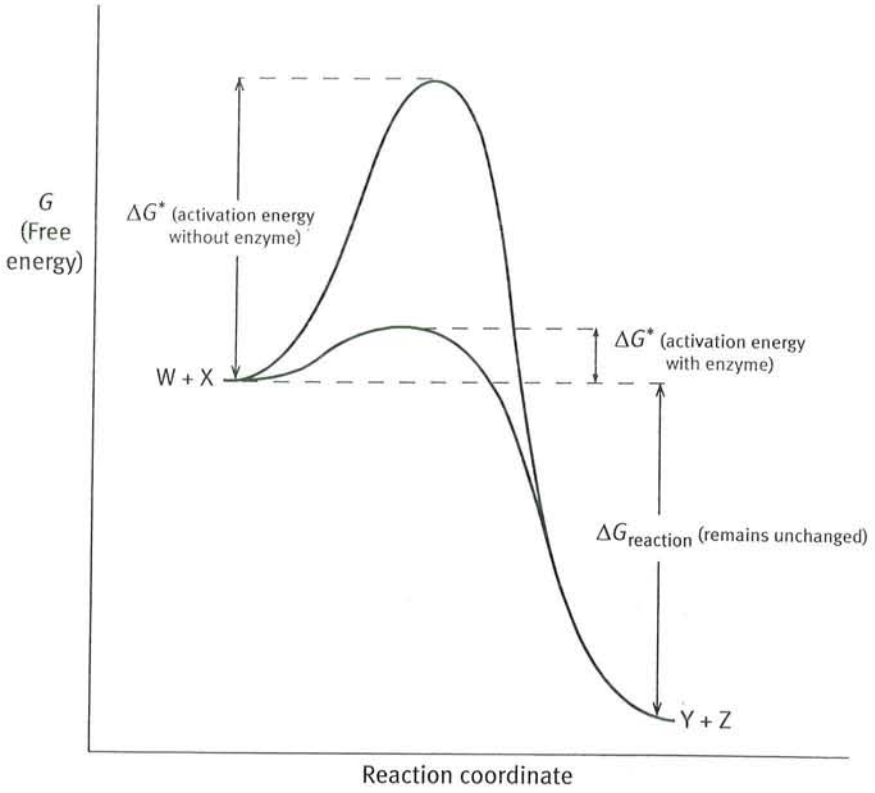


Figure 8.10. The Gibbs energies of activation, ΔG^{*}, of an uncatalysed and an idealized enzyme-catalysed reaction. The overall free energy change ΔG of the reaction is not altered by the presence of the enzyme. The reaction is spontaneous and will proceed at a faster rate at the same temperature in the presence of the enzyme.

few of the W –X complexes that happen to form will possess this amount of energy, so the reaction rate to form Y + Z will be slow (even though $\Delta G_{\text{reaction}}$ may be negative). In aqueous solution, a significant amount of the activation energy can be required to remove the hydration shells of W and X. Also there will be energy involved in rearrangement(s) of electronic charge within W and/or X during formation of the transition state. The reaction profiles of an uncatalysed and a simple idealized enzyme-catalyzed reaction (Figure 8.10) show that the enzyme has the effect of lowering the activation energy ΔG^{*} so that more W and X molecules achieve the transition state and the reaction will proceed at a faster rate.

To be useful, the study of enzyme activities must be quantitative. Any proposed mechanisms must be consistent with the experimental results. This is the province of the enzyme kinetics, a topic that concerns the rate at which enzyme-catalyzed reactions occur, and which provides information about enzyme mechanisms, inhibition, and regulation of activity. The Michaelis–Menten model of enzyme kinetics was introduced in 1913 and, despite many modifications, it is still the basic model for many enzymes.

$$\mathrm{E + S} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{ES} \xrightarrow{k_2} \mathrm{E + P}$$

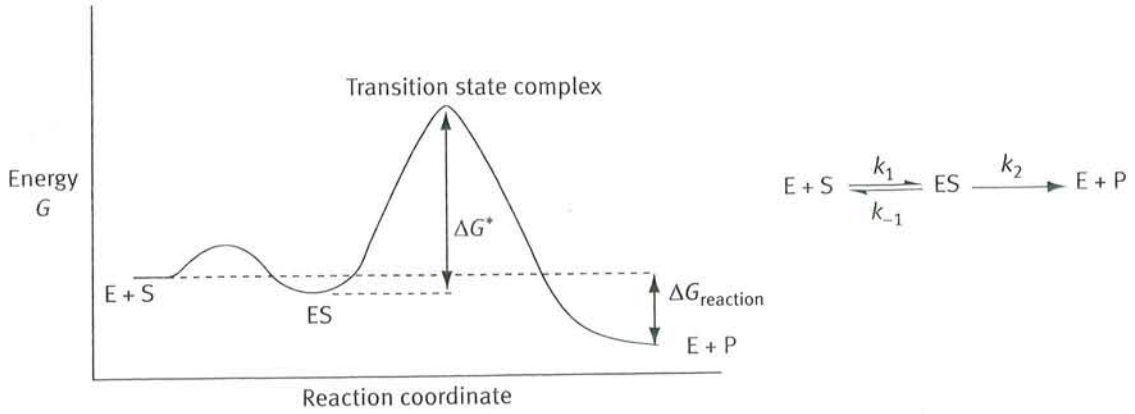


Figure 8.11. Reaction coordinate for a simple enzyme-catalysed reaction.

Enzyme (E) reacts reversibly with substrate (S) to form an enzyme–substrate complex (ES). In a second step, ES is converted to product (P) with release of the enzyme E. The rate constants (k) for the proposed stages are also shown.

Assumptions of the model:

1) There is no appreciable conversion of P back to S.
2) The concentration of ES quickly reaches a steady state, that is [ES] soon remains constant.

A reaction profile shows the energy levels at each step and provides a useful visual representation of the process. In kinetic terms, ΔG^* is the rate-limiting step (Figure 8.11).

In this profile the potential energy of the ES complex is lower than that of E + S. This is reasonable as ES is proposed to form spontaneously and rapidly. The process E + S $\longrightarrow$ ES will have a negative ΔG. The above figure gives a more complete description than that in Figure 8.10, but it is still far from describing what happens in a real enzyme-catalyzed reaction. Current ideas propose a much more dynamic process, with multiple transition states. Many conformational transitions take place and modification of the enzyme protein quite distant from the catalytic site can influence the conformational changes. Such results suggest that motions of the entire enzyme molecule are essential for catalysis and reinforce the idea that only 'adaptable' macromolecules are capable of producing the observed rates of reaction (Boehr *et al.* 2006; Benkovic *et al.* 2008; Hammes 2008).

The Michaelis–Menten equation can be expressed in mathematical form, which allows the experimental determination of the constants used to describe the enzyme activity in quantitative terms. A plot of rate of reaction vs substrate concentration for an enzyme-catalyzed reaction that follows Michaelis–Menten kinetics is hyperbolic in shape (Figure 8.12).

As the substrate concentration [S] is increased, the rate increase levels off, until it reaches the maximum achievable rate (V_{max}) for the conditions. Any further increase in [S] has no effect on the rate, as the enzyme active sites

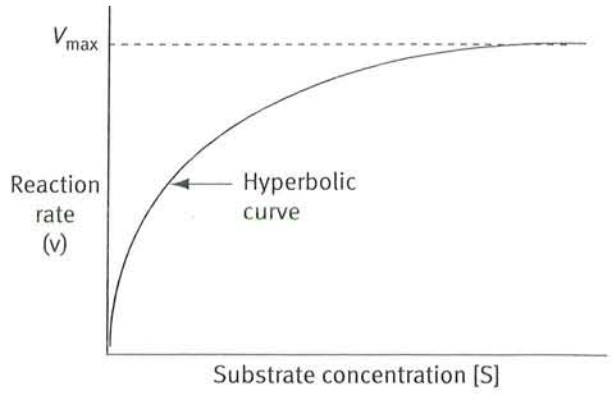


Figure 8.12. A plot of reaction rate vs substrate concentration for a reaction that follows Michaelis–Menten kinetics. The rate increase falls away as the substrate concentration approaches the saturation level of the enzyme. At saturation point of the enzyme with substrate, the maximum rate, V_{max}, is achieved.

are saturated. The V_{max} is an important kinetic characteristic of the enzyme under the given conditions. Enzyme kinetic data are now analyzed quickly using suitable computer programs. The Michaelis–Menten approach is a model, and like all models has its limitations. It doesn't apply to allosteric enzymes (see below). Several criteria and provisos must be considered when evaluating results from this and other models of enzyme kinetics, but I don't intend to pursue them here. If we imagine what is happening during an enzyme-catalyzed reaction, it would be something like this:

1) The reactants, substrates, of the enzyme usually attach quite specifically to the binding site, which positions them close to the active site. The active site is the region on the enzyme where the substrates actually undergo whatever bond-breaking and/or bond-making processes are involved in the reaction. This specific binding positions the substrates in an optimal orientation (the near-activity conformation (NAC)) for the formation of the transition state. The optimal orientation and associated proximity of the substrates substantially increases the probability of forming the required collision complex. NAC formation has been shown to occur from 1 to 7% of the time in active sites compared with 0.0001% for the uncatalyzed reaction (Bruice 2002). This effect contributes significantly to the rate of the reaction.
2) The binding of the substrates to the active site removes their solvation shells and leads to the exclusion of water from the site. Water is often a competing reagent and can alter the desired course of a reaction. This is important for many reactions. Release of 'bound' water molecules increases their mobility, often leading to a favourable increase in overall entropy.
3) As a result of interactions between the substrate and the amino-acid side-chains of the enzyme, there is an actual modification of the shape of the enzyme active site, so that the protein and substrate geometries are more complementary; they fit together better. The enzyme also forces the substrate into a conformation that approaches that of the transition state.

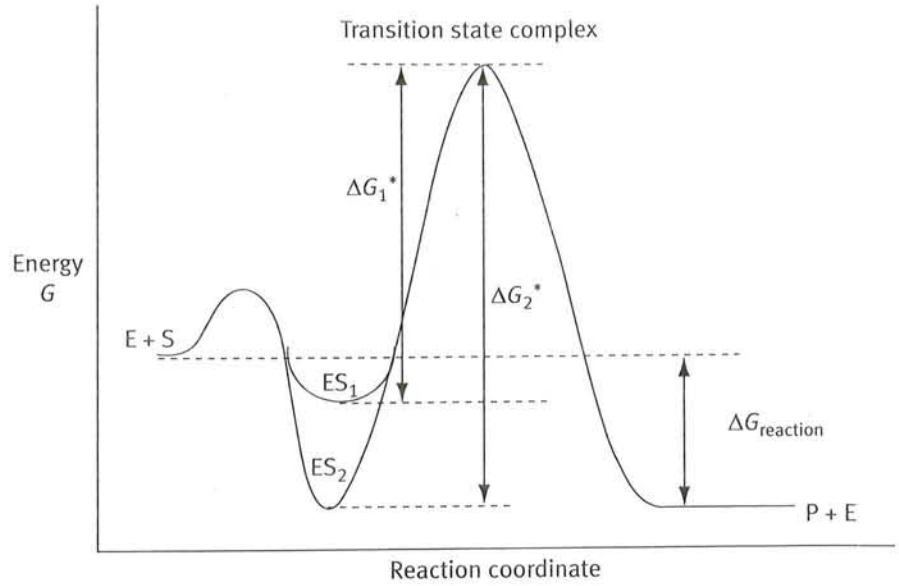


Figure 8.13. The case against tight binding of the substrate. ES_1 = the energy level of the ES complex for a typical enzyme-catalysed reaction. ES_2 = the energy level of the ES complex for a tightly bound substrate. As $\Delta G_2^* >> \Delta G_1^*$, rate 2 << rate 1.

This proposed dynamic adjustment of protein and substrate is part of the induced-fit hypothesis. An important result is stabilization of the transition state. The stabilization results in a great reduction in the activation energy needed to form the transition state, and leads to an increase in the rate of the reaction being catalyzed.

Although it is difficult to estimate quantitatively the relative contributions of the three factors mentioned above, it is generally agreed that the stabilization of the transition state is the most important. It is not the tight binding of the substrate. Very tight substrate binding would lower the energy level of the ES complex, thereby increasing the magnitude of ΔG^* and slowing the reaction (Figure 8.13).

The transition state stabilization idea is supported by the observation that so many enzymes have very high affinity for analogues of the transition state. A transition state analogue is a molecule that is very similar in structure to the transition state that forms with the natural substrate. However, in contrast to the real transition state, which is unstable and decomposes to form the products, the transition state analogue is a stable compound, and if well designed has the effect of blocking the activity of the enzyme. This is called inhibition of the enzyme. Many transition state analogues have been synthesized by chemists for the purpose of studying enzyme mechanisms and their inhibition. A practical application of such analogues is their use as drugs (see below). The study of inhibitors is important in enzymology. Inhibitors can be classified as reversible or irreversible. Reversible inhibitors fall into two main categories: competitive and non-competitive.

Competitive inhibitors involve a molecule similar in shape to the substrate, which binds to the substrate binding site and inhibits the binding of the natural substrate (Figure 8.14).

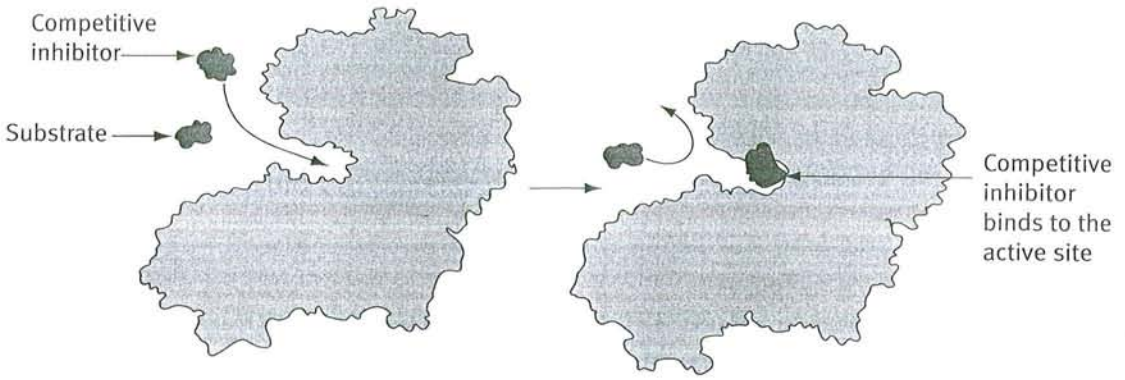
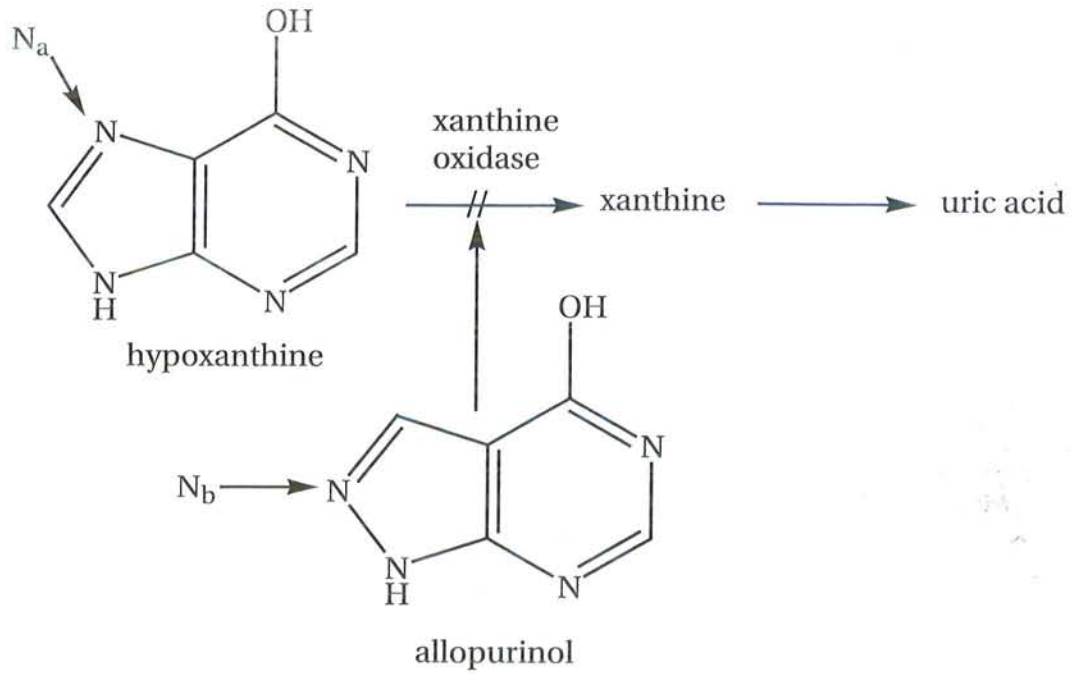


Figure 8.14. Competitive inhibition. A molecule similar in shape to the natural substrate competes with it for the active site. (Courtesy of Professor T. Steitz.)

A characteristic of competitive inhibition is that increasing the concentration of substrate will increase the likelihood of formation of an ES complex and increase the reaction rate in the presence of inhibitor. Allopurinol is a competitive inhibitor that reduces the production of uric acid in those who suffer from gout. It inhibits xanthine oxidase. Note the change in position of the nitrogen N_a in hypoxanthine to N_b in allopurinol (see below). The allopurinol is very similar in shape to hypoxanthine and competes with it for the binding site. The nitrogen atom N_b in allopurinol takes the place of a CH in hypoxanthine, so allopurinol cannot be oxidized to xanthine.

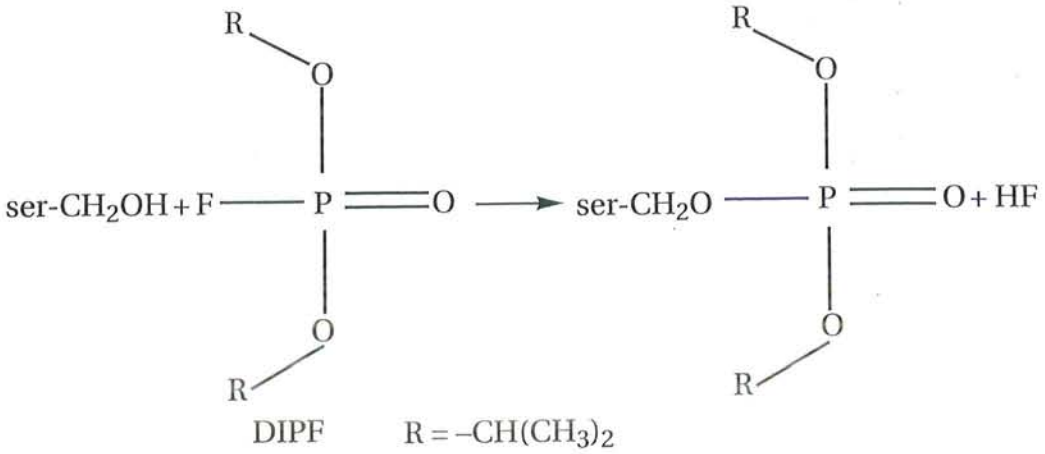


Drugs called statins reduce the rate of synthesis of cholesterol; Atorvastatin (Lipitor; Pfizer) is a transition-state analogue that inhibits the enzyme HMG-CoA reductase, a control-point enzyme in the cholesterol biosynthetic pathway.

The influenza drug Zanamivir (Relenza; GlaxoSmithKline) is a transition state analogue of the viral enzyme neuraminidase. Neuraminidase assists the invading virus to attach to host cells, so its inhibition reduces viral infectivity (von Itzstein *et al.* 1993).

Non-competitive inhibition involves interaction of the inhibitor with both enzyme E and the ES complex. The inhibitor is therefore not binding at the same place as S, so the inhibition cannot be reduced by increasing the concentration of substrate.

Irreversible inhibition occurs when the inhibitor attaches covalently to the enzyme and cannot be removed by dialysis. There is usually a time-dependent decrease in enzyme activity and the kinetic pattern resembles that of non-competitive inhibition. During studies of the group of enzymes called the serine proteases, a diagnostic test for the presence of serine in the active site was used. Serine has the side-chain CH_2OH, which reacts covalently with diisopropylphosphofluoridate (DIPF) and the enzyme is irreversibly inhibited.



DIPF $\quad R = -CH(CH_3)_2$

The penicillin antibiotics covalently attach to the serine group in the active site of glycoprotein peptidase, an enzyme that is involved in the synthesis of the bacterial cell wall. When this enzyme is inhibited, the cell wall is not fully crosslinked; it is easily ruptured and bacterial growth is stopped.

With the exception of irreversible inhibition, the attachment of the inhibitor to the enzyme depends on binding energy supplied by secondary forces, such as hydrogen, ionic, and van der Waals bonds, plus hydrophobic effects. The importance to biological activity of many individually weak bonds, precisely located in space, cannot be over-emphasized.

How does an enzyme actually work at the level of a chemical reaction? I shall use chymotrypsin as an example of the mode of action of an enzyme. By mode of action, I mean details of the way in which the enzyme catalyses the reaction. Chymotrypsin is a digestive enzyme that breaks certain peptide linkages in proteins, thus helping to degrade them in the small intestine. Enzymes that act in this way on proteins are known collectively as proteases. Chymotrypsin has been well studied for a number of years and its mode of action is known in detail. Chymotrypsin originates in the pancreas, but not as the active enzyme. If it were active there, it would do damage to proteins in the pancreas, as its 'job' is to degrade proteins. Instead, the pancreas produces an inactive precursor of

chymotrypsin, called chymotrypsinogen. Just as allosteric interactions can be used to alter the tertiary structure (and hence activity) of proteins, so the precursor of an enzyme can be cleaved and the tertiary structure altered. This transforms the precursor into an active enzyme. The inactive precursors are collectively known as zymogens.

The chymotrypsin zymogen, chymotrypsinogen, consists of a single polypeptide chain and has five disulphide (–S–S–) bonds which help determine its tertiary structure and keep it inactive as an enzyme. When chymotrypsinogen is released into the small intestine another protease, trypsin, cleaves it quite specifically between arginine-15 and isoleucine-16, forming π-chymotrypsin. The π-chymotrypsin then acts on itself to cleave out two pieces. This leaves the final form of the enzyme, α-chymotrypsin, which because of the cleavages of the original chain now has three shorter polypeptide chains held together by disulphide bonds. The cleavages bring about changes in the tertiary structure, so the final form is an active enzyme.

Chymotrypsin cleaves peptide bonds preferentially at amino acids which have aromatic (non-polar) side-chains, such as phenylalanine, tyrosine, and tryptophan. The selectivity for these side-chains is due to the size and hydrophobic nature of the side-chains lining the binding pocket of chymotrypsin. Precisely how does chymotrypsin achieve this cleavage of bonds, i.e. what is its mode of action? There are many ways in which enzymes catalyze reactions, but as a result of careful, detailed studies of hundreds of mechanisms, general principles applicable to all types have been recognized. In many enzymes the active site is located in a kind of pocket in the surface. A number of amino acids residues line the interior of the active site, and although not all are directly involved in the catalysis, they play a role in positioning the substrate or maintaining a water-free environment, should that be necessary.

Chymotrypsin is one of a group of enzymes called the serine proteases because the amino acid serine is located in the active site and is an essential residue involved in the catalysis. By means of studies making use of various chemical 'probes' it has been shown that serine-195 and histidine-57 are essential for the activity of chymotrypsin. These two residues must therefore be located close together in space. The amino acid sequence (primary structure) certainly doesn't give a clue to this. A look at their numbering will show that they are 138 residues apart! Obviously there has been a lot of folding of the chain to bring them into close proximity in the active site. This emphasizes the importance of the enzyme being folded in exactly the right way into its final conformation. The final conformation is maintained, as mentioned above, mainly by the cooperative energy of a large number of secondary bonds. The above conclusions based on chemical probes have been confirmed by means of X-ray crystallography, a powerful technique used to determine the three-dimensional structure of the entire enzyme. The chymotrypsin 'backbone' is folded mostly in pleated-sheet arrays and positions the essential residues precisely in the active site pocket. The role of all 240-odd amino acid residues is basically to act as a framework to achieve this result. What sublime design!

Chymotrypsin cleaves peptide bonds on the carboxyl side of aromatic amino acid residues. How does the enzyme 'know' how to do this? There are good

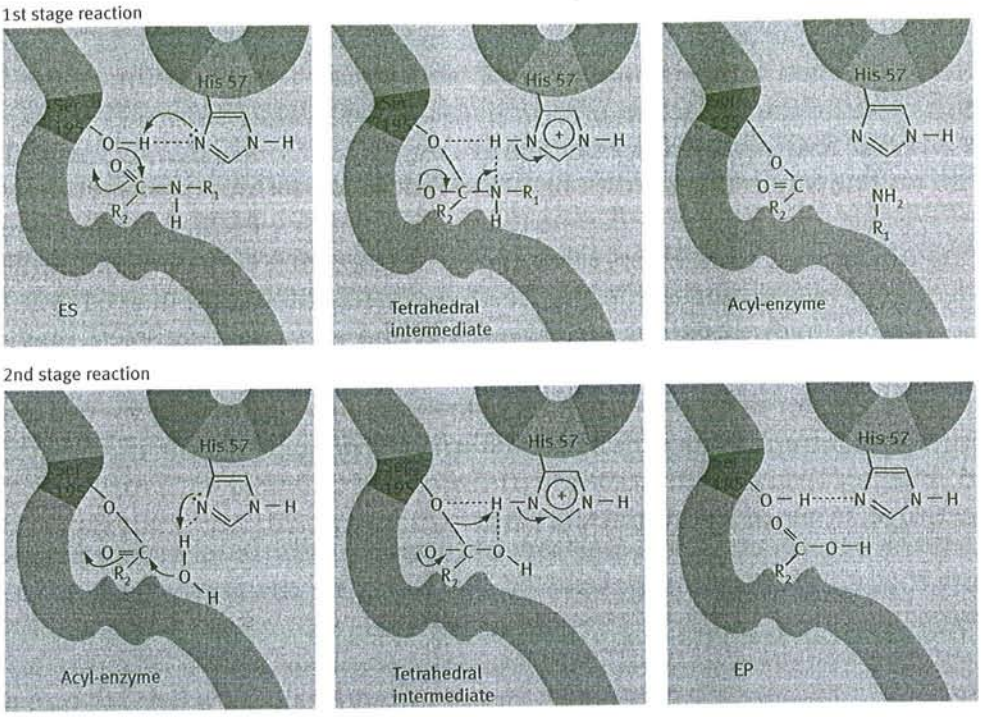


Figure 8.15. Mechanism of chymotrypsin action on a protein chain. (Illustration from Hammes, G.G. *Enzyme Catalysis and Regulation.* © 1982 Academic Press.)

structural reasons, leading to mutual recognition and interaction between enzyme and substrate, such that the peptide bond to be cleaved is in the correct position and orientation for catalysis. This is another example of de Duve's molecular complementarity at work (Figure 8.15).

The hydrolysis of the peptide linkage can be considered as taking place in two stages:

1) In an example of a nucleophilic reaction mechanism, the lone pair of electrons in the −OH group of serine-195 'attack' the partial positive charge on the carbon of the −C=O group in the peptide linkage. This leads, by a rearrangement of bonds as shown, to the acyl enzyme, where one piece of the cleaved substrate is covalently attached to serine-195. This demonstrates why the serine is essential to the activity of chymotrypsin and other serine proteases. If the serine −OH is blocked by a chemical probe specific for −OH groups, all enzyme activity is lost.
2) The acyl enzyme is attacked by the electrons on the oxygen atom of a water molecule, hence the formal term hydrolysis, or cleavage by water, is given to this type of reaction. After more bond rearrangements, as shown in Figure 8.15, both parts of the cleaved peptide linkage are released. One has a −COOH end group and the other an $-NH_2$ end group. These were the two groups formerly involved in the peptide linkage prior to cleavage. The serine-195 −OH group is restored, leaving the enzyme ready for the next substrate molecule. Note that histidine-57 was involved in both stages of the reaction. The lone pair on the imidazole nitrogen is protonated/deprotonated during the process.

Later research has revealed further details of the chymotrypsin mechanism. Aspartic acid-102 is also involved. Its γ-carboxylate group ($\gamma-COO^-$) is located close to a nitrogen atom in histidine-57 and assists in the protonation/deprotonation steps. Otherwise, the mechanism is the same as shown.

The overall hydrolysis reaction has a negative ΔG, which does not change whether an enzyme is involved or not. The thermodynamics of a reaction are not altered by the presence of a catalyst, but the kinetics most certainly are. The lowering of the activation energy barrier by the enzyme, by a combination of the several effects mentioned earlier in the chapter, has the result of greatly increasing the rate of hydrolysis. Strictly speaking, the enzyme-catalyzed reaction is achieved by means of a new mechanism.

How much faster rates are enzymes able to achieve, relative to an uncatalysed reaction? The rate enhancement varies greatly depending on the enzyme involved, but is commonly thousands of times the uncatalysed rate. Let's look at the decomposition of hydrogen peroxide, with and without catalysis:

$$2H_2O_2 \longrightarrow 2H_2O + O_2$$

Let the uncatalyzed rate = 1:

Conditions	Relative rate	Activation energy (kJ mol^{-1})
Uncatalysed rate	1	75
Platinum catalyst	2.8×10^4	49
Catalase (enzyme)	6.5×10^8	23

Platinum is commonly used in chemistry and catalyses the decomposition of H_2O_2 almost 30 thousand times faster than no catalyst. Catalase does the job a staggering 650 million times faster! This example is exceptional, as catalase has one of the highest turnover rates of any enzyme. Hydrogen peroxide and other highly active oxygen derivatives are intermediates in a variety of human disorders. Catalase is the predominant H_2O_2 – degrading enzyme in human erythrocytes. The importance of the protective role of catalase is reflected in its very fast catalytic rate (Meuller *et al.* 1997).Catalase is a good example of the enormous effect that lowering the activation energy from 75 to 23 kJ mol^{-1} has on the rate.

This doesn't seem such a great reduction for such an increase in rate, but the reason for the drop in rate becomes clear when the full calculation is made. The relationship between the rate constant of a reaction and the activation energy is given by the Arrhenius equation:

$$k = Ae^{-\Delta G^*/RT}$$

where k is the rate constant and A is a constant for the particular reaction concerned. From this we can see that k is inversely proportional to $e^{\Delta G^*/RT}$, so as ΔG^* increases, k decreases exponentially and the reaction rate will fall dramatically.

The above discussion of protein enzymes brings us to the point when we can answer an interesting and important question. Why are enzymes so large and complex?

Many people, including some chemists, are puzzled and amazed at the complexity of enzymes. Why should all that energy and information go into making a catalyst? I have mentioned before that evolution is parsimonious, especially in the expenditure of energy, so why spend so much? Chemists have devised some very successful catalysts, but such catalysts are incredibly simple and less sophisticated than enzymes. The answer to our question has been provided by research over recent years. To achieve the rate increases observed, all the subtleties of protein structure and movement have been harnessed by aeons of natural selection. Even parts of an enzyme molecule quite distant from the active site or binding site have been recruited to reduce the activation energy. The Arrhenius relationship shows that a linear reduction in activation energy results in an exponential increase in reaction rate, so every little bit helps. Also, some of the most fundamental electron-transfer reactions take place over what chemists consider to be large distances. Examples include the redox reactions of the electron-transport chain and photosynthesis. Electron transfer over distances of the order of 2 nm are involved. These can only occur because special channels exist between the electron donors and electron receivers in the enzymes involved. The dimensions and chemical nature of the channels are crucial, and need to be supported and surrounded by other parts of the protein. Although free water is excluded from the channels, in some cases particular hydrogen bonds are important in the electron-transfer processes. It appears that the parsimony biologists have noted at the organism level is maintained at the molecular level. Attempts by chemists and biochemists to 'clip' pieces from enzymes and reduce them to the bare minimum have usually shown that not much can be removed. A molecular weight of about 10,000 Da has been estimated as the lowest size likely for an active enzyme (Hammes 2008).

8.3 Regulation of enzyme activity

All the different modes of inhibition and activation may be demonstrated and studied quantitatively by various mathematical approaches to enzyme kinetics. Different models need to be applied to enzyme kinetics because it is important to emphasize just that—they are models of reality and simplifying assumptions often need to be applied. It is necessary for scientists to be aware of the assumptions and of the limitations of each model. Models are a vital part of science. An initial model may be proposed, based on the best available evidence at the time. As more results and/or better interpretations arise, the model may be modified, until it becomes as refined as possible. Models are seldom perfect, as they often attempt to deal with quite complex phenomena that are not always amenable to precise mathematical analysis or description.

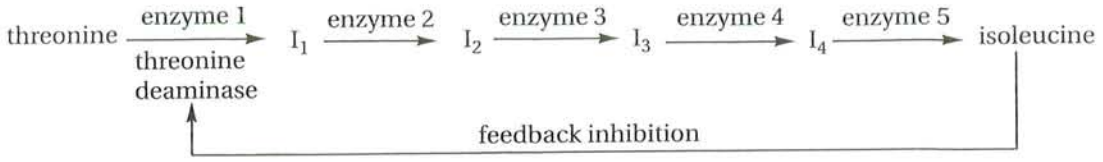
Enzyme catalysis of the thousands of reactions in a living cell cannot be haphazard. Synthesis vs degradation, energy storage vs energy use, provision of substrates here, removal of products there, all require careful regulation and

constant monitoring. Much of the monitoring is done by regulatory enzymes that sense the instant-by-instant needs of the cell and adjust their activity in response. We have spoken of enzymes in isolation, but most enzymes in living organisms are part of a metabolic pathway. Metabolic pathways need various controls or regulatory mechanisms to modulate their activity and maintain their metabolites at optimal levels. Regulatory mechanisms in specific pathways are dealt with in some detail in Chapter 10, so I'll give just an overview of the principles here. There are two rather obvious requirements that depend on factors outside the actual metabolic pathway under discussion. These are availability of precursors and availablity of coenzymes. The former is often the 'responsibility' of other metabolic pathways that feed in to the one under discussion. Not all enzymes require coenzymes, but those which do obviously cannot work without an adequate supply. A good example is that of the coenzyme couple NAD/NADH. At high levels of cell activity, the reduced form, NADH, is channelled into the respiratory chain from the citric acid cycle. High activity depletes the supply of the oxidized form, NAD, and this could be limiting if the respiratory chain for some reason failed to process the NADH back to NAD.

The activity levels of the enzymes of a metabolic pathway are crucial. These activities are controlled at three independent levels.

1) The expression of the enzyme protein at the level of the gene, that is the actual number of enzyme molecules being produced at a given time. The control here is exerted over the amount of messenger RNA coding for the enzyme protein and is termed transcriptional control. Transcriptional control involves hormones or metabolites that act on the gene control mechanisms. Increasing the levels of an enzyme in this way is called induction, whilst decreasing of the enzyme levels is called repression.
2) Enzyme interconversion, where the enzyme is already present, but in an inactive form, as in the case of chymotrypsin. When needed, the active form is produced by an activating enzyme. The activating enzyme often acts in response to a hormone signal, such activation being one of the roles of hormones. Some interconversions involve ATP-dependent phosphorylation of the enzyme involved. Protein kinases are enzymes that catalyze phosphorylation of serine, threonine, or tyrosine hydroxyl groups in proteins. The introduction of a doubly charged phosphoryl group can have large effects on protein functionality, and serves as a means of regulation. Interconversion control is more rapid than transcriptional control, as the enzyme (or its immediate precursor) is already present and ready for rapid modification. Modifications other than phosphorylation include carboxylation, acetylation, and glycosylation (addition of a sugar).
3) Regulation of key enzymes by ligands, which are small molecules that bind to allosteric proteins and change their conformation, as we have seen for haemoglobin. Allosteric enzymes are a major means by which the cell regulates enzyme activity to suit changing requirements. A ligand may be the

immediate product of the reaction or the end product of a metabolic pathway. The process is termed feedback inhibition. An example occurs in the biosynthesis of the amino acid isoleucine. Isoleucine is synthesized from threonine via five enzyme-catalyzed steps:



The first step is catalyzed by threonine deaminase. When the pathway is active, isoleucine accumulates and binds to its allosteric site on threonine deaminase, eventually stopping its own synthesis and saving cellular resources.

The feedback ligand might also be a metabolite from an entirely different pathway, but one intimately linked to the pathway under discussion. Depending on the requirements of the cell or tissue, both the levels and the activities of enzymes may be regulated to keep metabolic pathways running optimally.

Allosteric enzymes don't follow Michaelis–Menten kinetics. The rate vs substrate curve is sigmoidal rather than hyperbolic (Figure 8.16).

A sigmoidal curve is characteristic of cooperativity between subunits of the enzyme to bind the substrate and is a feature of allosteric enzyme kinetics. It means that binding of one substrate molecule to the first subunit makes it easier for additional substrate molecules to bind. Allosteric enzymes usually consist of two or more subunits. Each subunit has a substrate binding site and an allosteric ligand-binding site. The regulatory effect may be positive or negative, that is the rate can be increased or decreased relative to the unregulated rate (Figure 8.17).

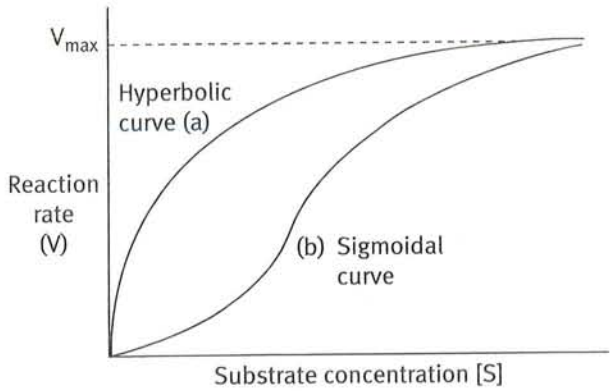


Figure 8.16. Reaction rate vs substrate concentration curves: (a) for an enzyme-catalysed reaction that follows Michaelis–Menten kinetics; (b) for an allosteric enzyme. The hyperbolic curve reflects the cooperativity characteristics of allosteric regulation.

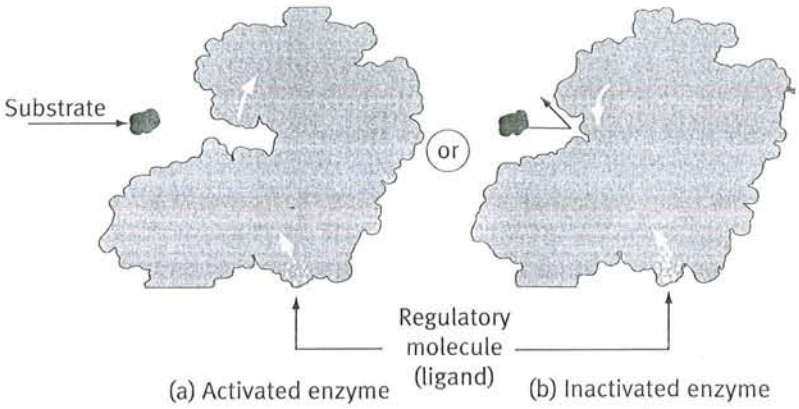


Figure 8.17. Allosteric control of enzyme activity. The regulatory molecule or ligand binds to a site distant from the active site. A regulatory molecule can (a) activate or (b) inactivate the enzyme by changing its conformation, and hence the access of substrate to the active site. Most allosteric enzymes consist of two or more subunits that act cooperatively. (Courtesy of Professor T. Steitz.)

8.4 Coenzymes, vitamins, and enzyme classification

For a substantial number of biochemical reactions coenzymes are required, so a supply of coenzymes must be available to the cell. Why are coenzymes necessary?

While some enzymes carry out their functions relying on their protein structures only, others need the help of non-protein compounds called cofactors. The term cofactor itself is usually reserved for certain metal ions that are essential for enzyme activity. The other cofactor types are organic molecules referred to as coenzymes. Coenzymes are further classified into soluble coenzymes, which bind to the protein portion during a reaction, assist in the reaction, and are then released, and prosthetic groups, which remain tightly bound, sometimes covalently linked, to the protein portion. Coenzymes are so important in biological reactions that they warrant discussion in their own right. It is not possible to cover them all here, but the following examples should give the reader an appreciation of their essential biological roles.

Redox enzymes or oxidoreductases are enzymes that catalyze oxidation/reduction (redox) reactions (see the enzyme classification below). They all require coenzymes and can be of the soluble or prosthetic group type.

1) The pyridine nucleotides NAD^+ and $NADP^+$ are widely distributed and have been mentioned a number of times already, but without detailed discussion. They are both soluble and move about the cell in this form. A number of enzymes use them as coenzymes and the NAD^+/NADH (oxidized/reduced) and $NADP^+$/NADPH ratios are important in cells, as is discussed in Chapter 11. It is instructive to look at the mechanism involving the pyridine nucleotides (Figure 8.18).

The full dinucleotide structures of NAD^+ and $NADP^+$ shown in Figure 8.18(a) are partly replaced by R in Figure 8.18(b) to allow focus on the transfer of H and electrons. When NAD^+ is complexed with an enzyme and the appropriate substrate, there is net removal of two H atoms from the substrate (e.g. a foodstuff)

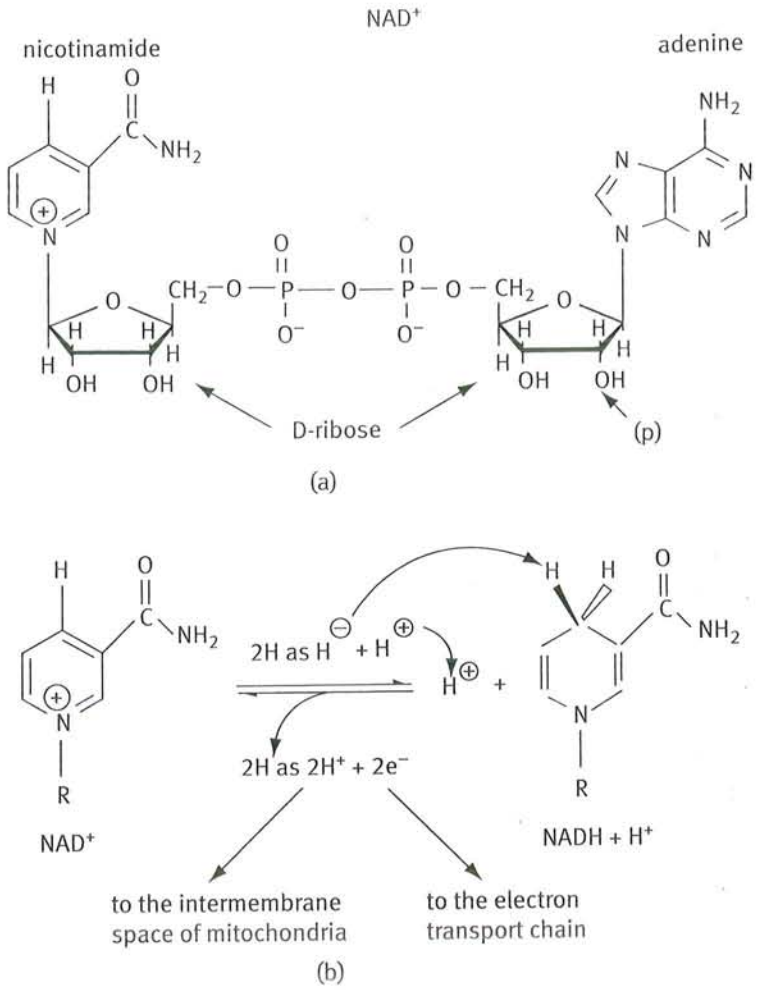


Figure 8.18. (a) The structure of NAD^+, nicotinamide adenine dinucleotide. (p) indicates the point of attachment of a phosphate group to form $NADP^+$. (b) The reduction of NAD^+ to $NAD + H^+$. Transfer of hydride ion H^- to nicotinamide is stereospecific. It always adds to the same 'face' of the ring. In the reverse direction the release is in the form of protons and electrons, the latter to special receiver molecules.

being oxidized. These H atoms are removed as hydride ion H^- and a proton H^+, as shown. In the reoxidation step, there is net removal of two the H atoms as two protons (pumped into the intermembrane space of the mitochondrion) and two electrons that are transferred along the ETC. The reoxidized NAD^+ is returned to the coenzyme pool for reuse. In Chapter 10 we discuss the roles of NAD^+/NADH, the coenzyme couple involved in many energy-producing (catabolic) oxidations in the cell. The role of $NADP^+$/NADPH is mainly in anabolic pathways (Chapter 11) that utilize energy to build complex molecules.

2) The flavin coenzymes flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD) have flavin as the redox group (Figure 8.19).

FMN has the same flavin moiety as FAD, but its side-chain R is less complex. In FMN, R is ribitol with one attached phosphate group at position 5 as shown in

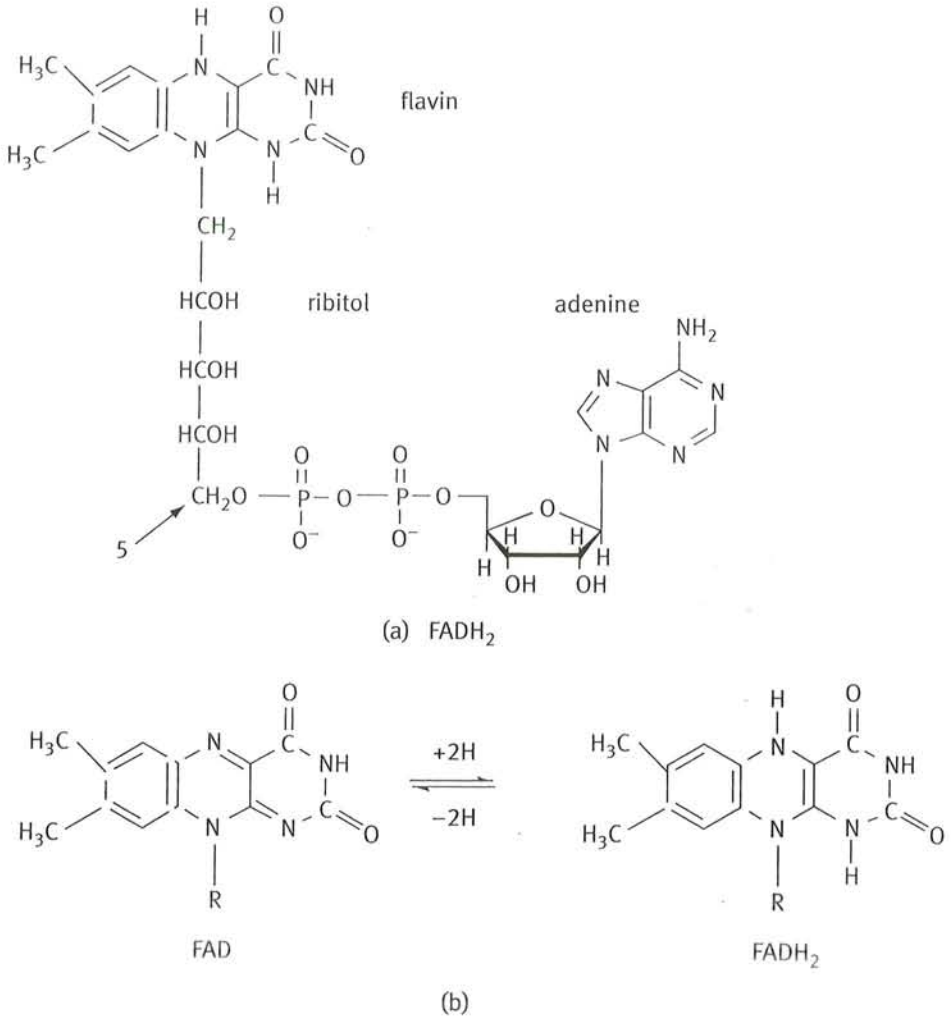


Figure 8.19. (a) The structure of $FADH_2$. (b) The flavin moiety of FAD undergoes reduction by two hydrogen atoms (as H^+ and H^-) from a substrate, such as succinate from the tricarboxylic acid cycle (see Chapter 11).

Figure 8.19(a). FAD and FMN are covalently linked to proteins as prosthetic groups, and form a group called flavoproteins. As well as transferring two hydrogen atoms, FAD and FMN can transfer one proton and one electron at a time when necessary. During this process, a radical intermediate is formed. Radicals are highly reactive and potentially damaging species, so FAD and FMN remain bound to the enzyme protein and are not allowed to be 'free' to move about the cell. FAD and FMN are found in dehydrogenases, oxidases, and monooxygenases.

Group-transferring coenzymes transfer entire groups of atoms as a single unit.

1) Pyridoxal phosphate is an important coenzyme in amino acid metabolism. Its role in transamination reactions is illustrated in Figure 8.20.

The pyridoxamine phosphate (Figure 8.20(a)) is involved in the transfer of $-NH_2$ groups from one amino acid to an oxoacid to form a new amino acid, catalyzed by a transaminase:

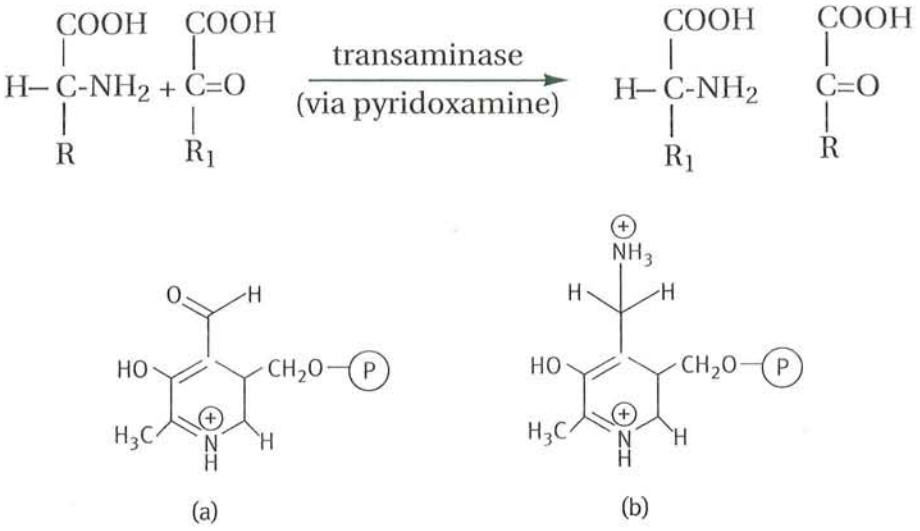


Figure 8.20. (a) Pyridoxal phosphate, aldehyde form. (b) Pyridoxamine phosphate, an intermediate in transamination reactions.

2) Biotin is the coenzyme used by carboxylases. Using ATP, biotin reacts with HCO_3^- to form *N*-carboxybiotin, from which CO_2 is then transferred to other molecules (Figure 8.21).

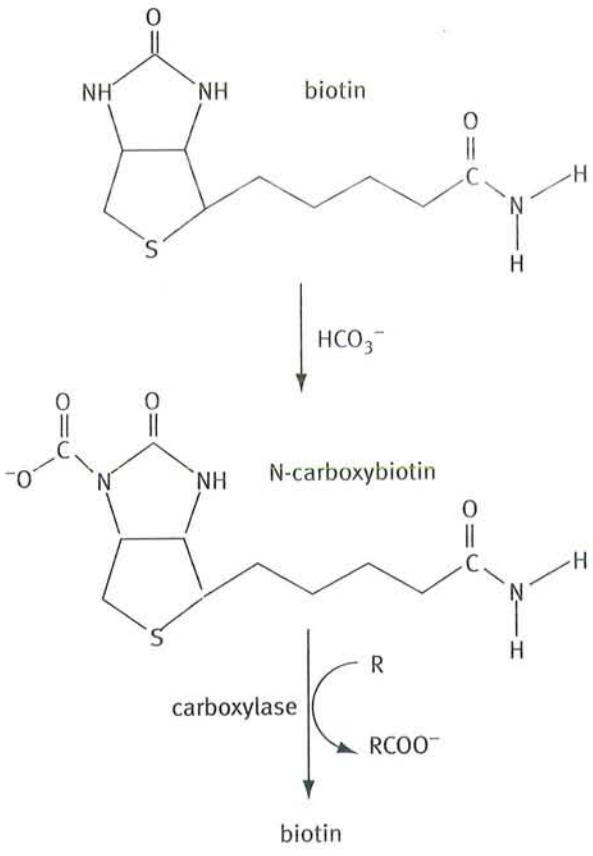


Figure 8.21. Biotin assists in the transfer of carbon dioxide units. Example: R = pyruvate, CH_3COCOO^-; $RCOO^-$ = oxaloacetate, $^-OOCCH_2COCOO^-$.

Many of the compounds we know as vitamins are the precursors of coenzymes. As vitamins are not made in the human body, but are essential for good health, they must be provided in the diet. Traditionally, the vitamins are divided into two groups: water and fat soluble. With the exception of vitamin C, the water-soluble vitamins are components of or precursors of coenzymes. The side-chains of amino acids that make up proteins have only a limited range of chemical activities. The association between the enzyme protein and a coenzyme extends the number of reaction types that can be carried out. The coenzyme is usually modified during the reaction and is later converted back to its original form by other enzymes (e.g. the NAD^+/NADH system). This allows small amounts of these precious substances to be used repeatedly, time after time.

Vitamin deficiencies due to insufficient intake in the diet are the cause of a number of diseases in humans. Considering their key roles in metabolism, this is not surprising. The story of the gradual realization that such compounds existed is long and rather complex, partly because the biochemistry of their action was not known at the time. Early researchers found that certain diseases could be cured by extracts from certain foods, but they had no idea why. The division into water- and fat-soluble arose when researchers tried to purify and identify these mysterious factors. The fat-soluble group could be dissolved in ether, and thus could be separated from the water-soluble group, which could not. Eventually, separation and purification techniques were developed that enabled the individual compounds to be fully identified. This was no easy task, as the chemical structures of the various vitamins vary widely. Table 8.1 summarizes some of the important aspects of the vitamins, the coenzymes, and their associated deficiency diseases.

Table 8.1. Coenzymes, their precursors, reactions catalysed, and deficiency diseases.

Coenzyme	Precursor	Reaction type	Deficiency disease
Biotin	Biotin	Carboxylation	Anorexia, dermatitis
Coenzyme A	Pantothenic acid	Acyl transfer	Dermatitis
FAD, FMN	Riboflavin (B_2)	Redox	Growth retardation
NAD^+, $NADP^+$	Niacin	Redox	Pellagra
Pyridoxal phosphate	Pyridoxine (B_6)	Transamination	Dermatitis
Thiamine pyrophosphate	Thiamine (B_1)	Aldehyde transfer	Beriberi

The fat-soluble vitamins are not directly related to coenzymes; they are involved in vision (vitamin A), maintenance of bone structure (vitamin D), blood coagulation (vitamin K), and antioxidant processes (vitamin E).

Let's look at two further aspects of enzymes that concern us all: deficiencies in the control of their expression and the use of enzymes as markers for disease.

Some diseases are caused by what are known as 'inborn errors of metabolism'. These are genetically transmitted diseases caused by abnormalities in the DNA of genes, resulting in diminished levels of a particular enzyme. An example is phenylketonuria (PKU). The name arises from the fact that a metabolite called a phenyl ketone accumulates and appears in the urine. A deficiency in the level of the enzyme phenylalanine carboxylase causes this accumulation. The problem caused is mental retardation, probably as a result of a build up of the phenyl

ketone and related metabolites in the brain. This results in an osmotic imbalance in which water flows into the cells, causing them to expand and crush one another in the developing brain. PKU is easily detected in newborn babies and a test for its presence is mandatory in the USA. The treatment is a modified diet, low in phenylalanine and enhanced in tyrosine (the missing product). Sufferers must avoid the artificial sweetener aspartame, which is a peptide containing phenylalanine. A substitute for aspartame, which contains alanine rather than phenylalanine, is available.

Some enzymes may be used as 'markers' or indicators for disease. Lactate dehydrogenase (LDH) has two isoforms that are slightly different in amino acid composition, so with a suitable test they are distinguishable. One form occurs in heart muscle and the other in skeletal muscle. Any increase in LDH levels in the blood indicates tissue damage of some kind. A heart attack can be diagnosed with a high degree of certainty by detecting increased levels of the heart-muscle LDH in the blood. Similarly, acetylcholinesterase, an enzyme involved in controlling certain types of nerve impulse, is a useful marker. A number of pesticides will interfere with this enzyme, so farm workers suspected of excessive exposure to pesticides may be tested for acetylcholinesterase activity. More than 20 enzymes are regularly used as clinical markers. They are used to examine the function of the liver, heart, brain, prostate gland, pancreas, and red blood cells, to name a few. All this is made possible by the depth of knowledge biochemists now have of so many enzymes and their unique distribution patterns in certain organs and tissues.

There is still a long way to go. Not all proteins believed to be encoded on our genes have been identified, let alone characterized. This identification task is being attempted by practitioners in the field of proteomics. The human genome projects have succeeded in identifying and sequencing essentially all our genes. Functional genomics aims to look at all the genes expressed during particular metabolic process or during embryonic development. It involves identifying the transcriptome, the set of messenger RNAs transcribed from DNA under a defined set of conditions.

Glycomics has been added to the growing 'omics' list recently. A large percentage of proteins are glycosylated, that is have carbohydrate attached. Glycomics seeks to identify the roles of all the carbohydrates that are attached to proteins and cell membranes. Metabolomics is immensely ambitious as it seeks to identify and quantify all the metabolites in a system or organism. Such ambitions have only become realistic with the availability of modern instrumentation and the expansion of computational capabilities. The task ahead is huge.

Just how many different types of chemical reaction can enzymes catalyze? Biochemists have naturally been vitally interested in this. Much work has been carried out on classifying enzymes into the minimum number of groups possible, based on the basic mechanism of the reaction catalyzed. Perhaps surprisingly, it has been possible to fit the large number of enzymes into just six major groups or classes. All enzymes, once fully characterized, are entered in the Enzyme Catalogue (EC) and given a four-digit EC number. The first digit indicates the major class, and the rest the subclass, sub-subclass, and position in the

sub-subclass. As an example, let's take alcohol dehydrogenase (ADH), which catalyses the reaction:

$$\text{ethanol} + \text{NAD}^+ \longrightarrow \text{acetaldehyde} + \text{NADH} + \text{H}^+$$

or

$$\text{CH}_3\text{CH}_2\text{OH} + \text{NAD}^+ \longrightarrow \text{CH}_3\text{CHO} + \text{NADH} + \text{H}^+$$

The EC number for ADH is 1.1.1.1.

The name comprises several levels:

Class 1. Oxidoreductases (catalyze oxidation–reduction reactions)

Subclass 1.1 $-$CHOH is the electron donor

Sub-subclass 1.1.1 NAD^+ is the electron acceptor

Its place in the sub-subclass is 1, so the whole EC number is 1.1.1.1[Zn^{2+}].

The fact that zinc, as the ion Zn^{2+}, is necessary for activity, is also indicated.

The six enzyme classes are:

1) oxidoreductases: catalyze the transfer of reducing equivalents between redox systems
2) transferases: catalyze the transfer of functional groups from one substrate to another
3) hydrolases: also involve group transfer, but the acceptor is always a water molecule
4) lyases (synthases): catalyze the removal or formation of a double bond
5) isomerases: catalyze the movement of groups from one place to another, within the same molecule
6) ligases (synthetases): catalyze the formation of C–C, C–O, C–N, or C–S bonds. They are energy dependent and are coupled to the hydrolysis of nucleoside triphosphates (e.g. ATP).

Enzymes having the same function are often found in a wide range of organisms. This is particularly the case with enzymes of the major metabolic pathways that are common to most forms of life. The study of the same enzyme across a number of species may be useful in a number of ways. Depending on how long ago two species may have diverged, there will be various changes in the amino acid sequence of an enzyme (or indeed any protein) that is common to both. These changes are due to mutations in the DNA of the genes, whose base sequences determine the protein amino acid sequence, as described in Chapter 9. Comparison of protein sequences in different species is used in studies of the relationship between amino acid sequence and protein structure. If the protein has maintained the same function in different species, despite having some sequence changes caused by mutation, then the mutations must not have altered the structure significantly, that is the protein function must have been 'tolerant' of such mutations. No doubt there could have been other mutations that did seriously affect the function, but the individuals carrying them would have been unlikely to survive. Protein chemists can thus use these protein homology studies to derive general principles of protein sequence/structure relationships.

This assists them in studies on the all-important three-dimensional structure of proteins and enzymes.

The number and types of mutations found in proteins can also provide information on the evolutionary relationship between species, as described for nucleic acids in Chapter 9. Until recently, it was quicker to sequence proteins than nucleic acids. One well-studied example is the pair of oxygen-binding haem proteins haemoglobin and myoglobin of humans, where there is considerable sequence homology, strongly suggesting evolutionary divergence from a common ancestral molecule. Another example, ranging across a number of species, involves the amino acid sequence of cytochrome *c*, a component of the respiratory chain (see Chapter 10). In fact, a phylogenetic tree, a diagram illustrating the evolutionary relationships between a group of organisms, has been constructed using the sequences of cytochrome *c* in another example of molecular evolution. The tree constructed on the basis of differences in primary structure of cytochrome *c* shows close agreement with results derived from more classical methods, such as the fossil record. These results support the idea that closely related proteins share a common evolutionary origin and provide further convincing evidence in favour of the theory of evolution in its modern form.

REFERENCES

Benkovic, S.J., Hammes, G.G., and Hammes-Schiffer, S. (2008) *Biochemistry* **48**, 3317–21.

Boehr, D.D., McElbeny, D., Dyson, H.J., and Wright, P.E. (2006) *Science* **313**, 1638–42.

Bruice, T. (2002) A view at the millennium: the efficiency of enzymatic catalysis. *Accounts of Chemical research* **35**, 139–48.

Garrett, R.H. and Grisham, C.M. (1999) *Biochemistry*, 2nd edn. Saunders College Publishing, Fort Worth, TX, p. 191.

Hammes, G.G. (1982) *Enzyme Catalysis and Regulation*. Academic Press, New York.

Hammes, G.G. (2008) How do enzymes really work? *Journal of Biology and Chemistry* **283**, 22337–46.

Luisi, P.L. (2006) *The Emergence of Life*. Cambridge University Press, Cambridge, pp. 245–46.

Meuller, S., Riedel, H.-D., and Stremmel, W. (1997) Direct evidence for catalase as the predominant H_2O_2 – removing enzyme in human erythrocytes. *Blood* **90**, 4973–8.

Riemer, J., Bulleid, N., and Herrmann, J.M. (2009) Disulfide formation in the ER and mitochondria. Two solutions to a common process. *Science* **324**, 1284–7.

von Itzstein, M., *et al.* (1993) Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature* **363**, 418–23.