2

An introduction to proteins and peptides

The fundamental component of a protein is the polypeptide chain composed of amino acid residues; twenty different residues are involved in protein synthesis. These residues might be modified after the synthesis of the polypeptide chain. The other components of proteins are called prosthetic groups. The structure of the amino acids and their characteristic property as amphoteric molecules is described, followed by a description of asymmetry and chirality. The way in which amino acid residues interact within proteins is explained. The ionic properties of proteins are important in such interactions and in their electrophoretic separation. Proteins can also be separated on the basis of their size. After mentioning how the order of the amino acid residues in polypeptides can be determined, the hierarchies of protein structure are briefly described. The tertiary structure of proteins can be destroyed by denaturation. Finally, it is shown that even small peptides can possess biological activity, for example as hormones and transmitters.

THE ROLE OF AMINO ACIDS IN THE CELL

Amino acids are a fine example of the versatile roles performed by the cell constituents. Amino acids contain, among other functional groups, two that are common to all amino acids: an amino (or imino) group and a carboxyl group. The ability of an amino acid to condense with other amino acids to form a peptide is dependent on the chemical properties of these two functional groups. Certainly, a most important role for amino acids is to serve as the monomeric subunits of proteins, but they have other important roles. For example, the tripeptide glutathione has an important function and other small peptides serve as hormones and, in some organisms, as antibiotics; glutamic acid acts as a neural transmitter. Amino acids are the precursors of a wide variety of biomolecules (e.g. nitric oxide from arginine, histamine from histidine). Some amino acids are metabolized and utilized for the production of glucose (gluconeogenesis). As there is no store of amino acids, apart from those involved in protein structure, proteins have to be broken down to free amino acids when the latter are required for gluconeogenesis.

STRUCTURE OF AMINO ACIDS

All the common amino acids, except for proline, have the same general structure in that the α-carbon atom bears a –COOH group, an $–NH_2$ group and an 'R'-group, which is responsible for the different properties of the various amino acids. A general formula for amino acids is shown in Fig. 2.1. The structures of the 20 common amino acids are shown in Fig. 2.2, grouped according to the nature of their R-groups. The internationally-approved three-letter and single-letter abbreviations for each amino acid are also indicated.

The α carbon is optically active in α-amino acids other than glycine. The two possible isomers are termed D and L. All naturally occurring amino acids found in proteins are of the L-configuration (see p. 9).

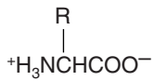
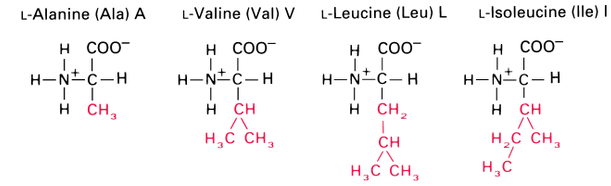
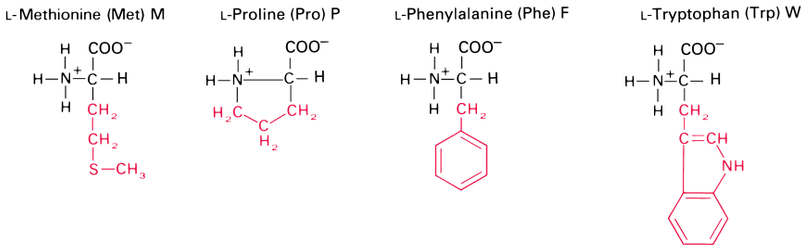


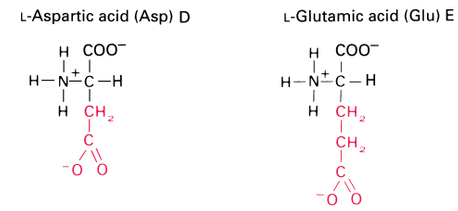
Fig. 2.1 General formula of an amino acid.

1. Non-polar or hydrophobic R-groups

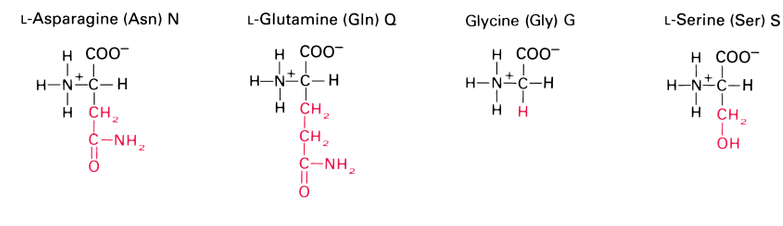


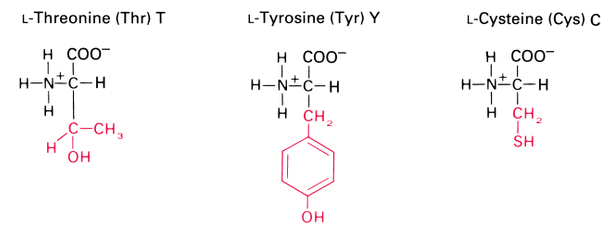


2. Negatively charged R-groups at pH 6–7



3. Uncharged or hydrophilic R-groups





4. Positively charged R-groups at pH 6–7

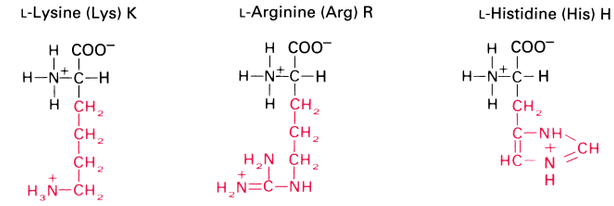


Fig. 2.2 Structures of the 20 common amino acids grouped according to the nature of their 'R'-group. Note the three- and one-letter notations.

A cystine residue is formed from two cysteines linked through a disulfide bridge (–S–S–) formed from their sulfhydryl (–SH) groups.

The charges on the amino acids indicated in Fig. 2.2 are those that occur at pH 6–7. Acids are defined as proton donors and bases as proton acceptors. It follows that, at pH 6–7, an amino acid in group 2 is present as a free base (an anion) and one in group 4 as a free acid (a cation). The terms 'acidic' and 'basic', as applied to amino acids, should therefore be used with caution because they refer to the protonated forms of group 2 or the unprotonated forms of group 4. A compound such as an amino acid that carries both basic and acidic groups is referred to as amphoteric.

ASYMMETRY IN BIOCHEMISTRY

ASYMMETRY AS APPLIED TO AMINO ACIDS AS AN EXAMPLE

Chirality is derived from the Greek word *cheir* for 'hand' – the left and right hands are mirror images of each other. Such asymmetry in molecular structure is of great importance in biochemistry. A chiral molecule possesses at least one asymmetric centre, such as a carbon atom, to which are joined four groups that are different from each other.

The amino acid alanine can exist in two forms, denoted D-alanine and L-alanine, as shown in Fig. 2.3. The amino acids contained in mammalian proteins are of the L-form. (Sugars are also chiral molecules; D-sugars predominate in mammalian carbohydrates; see p. 110.) In Fig. 2.3, red denotes the oxygen atoms of the carboxyl group, the nitrogen atom of the amino acid group is grey, the carbon atoms are black and the hydrogen atoms are white.

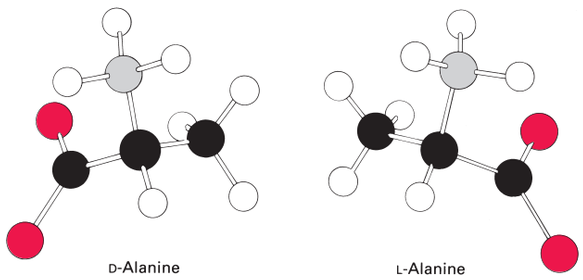


Fig. 2.3 D- and L-alanine are examples of asymmetry in structure.

NON-CHIRAL ASYMMETRY

Even if a molecule is not chiral, it can contain identical groups that are sterically distinguishable. A simplified representation of a hypothetical molecule is shown in Fig. 2.4. If A and B are held in space on a surface, then the identical groups X_1 and X_2 can be distinguished. The classic biochemical example is citric acid. Although this molecule has a plane of symmetry, the central carboxyl group and the hydroxyl group can be held in such a way that the two $-CH_2COOH$ groups can be distinguished and the molecule is able

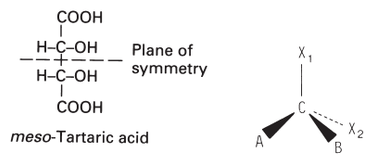


Fig. 2.4 Example of a *meso* compound (left) and the simple representation of a hypothetical molecule (right).

to interact with an enzyme that has specific binding sites for the different groups in the molecule (see Fig. 9.15, p. 136). Such a molecule is termed 'prochiral' in that it can be made chiral by changing the structure of the group on only one of the central carbon bonds. Note that, if a molecule has a plane of symmetry such that chiral centres on either side of the plane of symmetry exactly compensate, the molecule is termed a *meso* compound (e.g. *meso*-tartaric acid, shown in Fig. 2.4).

R AND *S* CONVENTION

A chiral centre can be denoted *R* or *S*. The method for ascribing the *R* or *S* designation to a centre is as follows:

- List the functional groups in order of their priority assigned by convention. The order for some biochemically important groups is –SH (highest), –OH, $-NH_2$, –COOH, –CHO, $-CH_3$, –H (lowest). Then orientate the molecule so that the group of lowest priority points away from the observer.
- If the order of priority (high to low) of the remaining groups is clockwise, the centre is *R*. If the order or priority is anticlockwise, the centre is *S*. Thus the α-carbon of L-alanine has the *S* configuration.

IONIC PROPERTIES OF AMINO ACIDS

ELECTROPHORETIC SEPARATION

As already explained, amino acids have amphoteric properties that allow their separation by electrophoresis at pH 6.0, in which the amino acids move along a medium (paper) under the force of an applied electric field. Such a separation is illustrated in Fig. 2.5. Electrophoresis is commonly carried out on paper but gels can also be used. The amphoteric nature of α-amino acids means that, in the absence

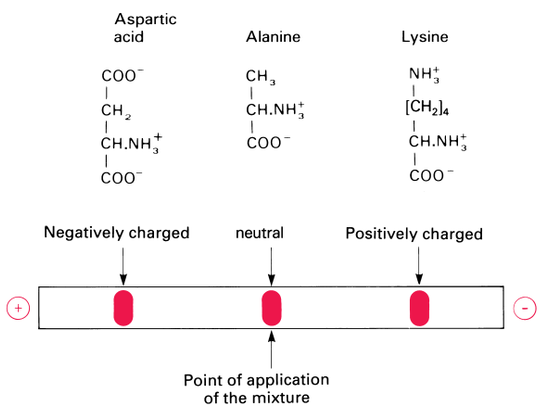


Fig. 2.5 Demonstration of the ionic properties of amino acids by electrophoresis.

of other acids or bases, the carboxyl and amino groups are both fully ionized, giving rise to the term zwitterion (German *Zwitter* = hybrid or hermaphrodite). This is the form that predominates in neutral solution and in crystals, rather than the unionized form.

THE BUFFERING CAPACITY OF AMINO ACIDS

As explained previously, an acid is defined as a proton donor. Acids vary in their tendency to dissociate; stronger acids do so more readily than weaker ones. The strength of an acid is expressed by the term pK_a, which is the pH at which an acid is 50% dissociated. The titration curve of alanine (Fig. 2.6) shows that the –COOH group becomes more dissociated as the pH increases. At pK_1, the change in pH of the solution with increasing additions of NaOH is lowest; in other words, the buffering capacity is greatest.
We have defined a base as a proton acceptor; so, in this case, the proportion of ionized NH_3 decreases as the pH increases and the maximum buffering is at pK_2.

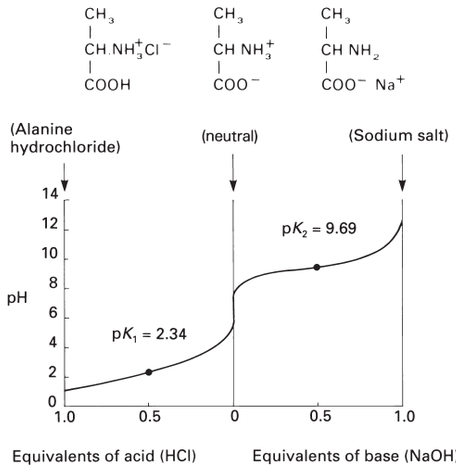


Fig. 2.6 Amino acids possess buffering capacity, as demonstrated by the titration of alanine.

The buffering capacity of histidine

If the R-group of an amino acid is capable of being ionized, then the amino acid will have a third p*K*. Histidine is very important in this respect because the imidazole group is only weakly basic, having a pK_a of 6.00. It therefore exists as a mixture of the protonated and dissociated forms in solution at the physiological pH of 7.2–7.4. Histidine therefore contributes to the buffering capacity of proteins. The titration curve of histidine is shown in Fig. 2.7.

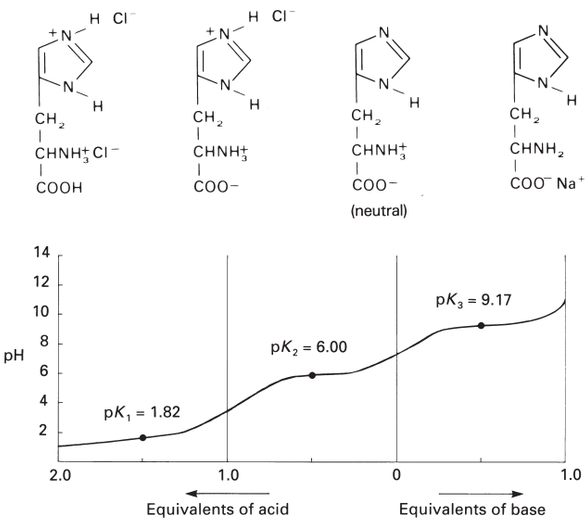


Fig. 2.7 The buffering properties of histidine are of particular physiological importance.

SEPARATION OF AMINO ACIDS BY ION-EXCHANGE CHROMATOGRAPHY

It is often important to determine the proportion of the different amino acids, either in body fluids such as serum or spinal fluid, or in a protein hydrolysate. For this purpose, a resin bearing either positively charged groups (anion-exchange resin) or negatively charged groups (cation-exchange resin) can be used. Amino acids passed down a column of such a resin bind competitively to the charged groups on the resin.

Figure 2.8 shows the separation of the amino acids present in a peptide hydrolysate on a column of sulfonated polystyrene (cation-exchange resin). Passage through the column of buffers of increasing pH causes aspartic acid (acidic) to emerge as the first amino acid and arginine (basic) as the last. It is common to detect the amino acids using ninhydrin; this gives a blue colour after reaction with all α-amino acids (yellow for the amino acid proline), the intensity of colour being related to the amount of the particular amino acid. The whole process can be automated.

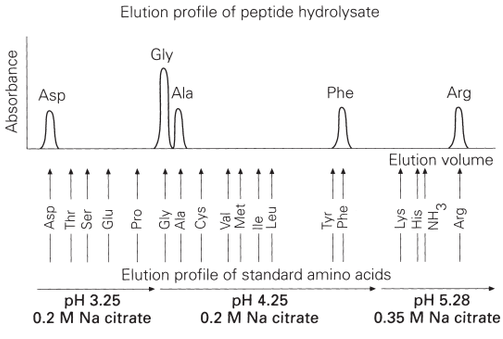


Fig. 2.8 Amino acids can be separated by ion-exchange chromatography and the amount of each determined.

Similar methods can be used for the separation of proteins that carry various net charges.

PEPTIDE STRUCTURE AND THE PEPTIDE BOND

THE PEPTIDE BOND

The peptide bond is formed by the interaction of two amino acids, with the elimination of water between the neighbouring $-NH_2$ and $-COOH$ groups. This is shown in Fig. 2.9. The peptide bond is a rigid structure; this has important implications for the structure of proteins (see p. 56).

Proline can also participate in a peptide bond (Fig. 2.10) but, in contrast to the α-amino acids, there is then no H available for H bonding which, as we will see, is important in the secondary structure of proteins.

NOTATION USED FOR PEPTIDES

The structure of a typical peptide, enkephalin, is shown in Fig. 2.11. In writing the primary structure, one starts with the amino-terminus (also called the N-terminus) and ends with the carboxy-terminus (referred to as the C-terminus). In Fig. 2.11, enkephalin is given in the three-letter code for amino acids. Abbreviated according to the single-letter code it would be written YGGFM.
A peptide composed of more than a few amino acid residues is termed a polypeptide. To the extent that such polypeptides are the backbone structures of proteins, there is no formal definition of a transition from polypeptide to protein, but insulin, which has 50 amino acid residues, is commonly regarded as being typical of the smallest protein.

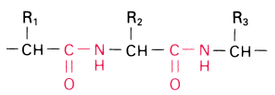


Fig. 2.9 Peptide bonds are formed by the interaction of amino acids.

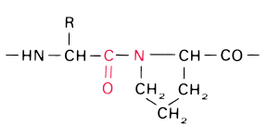


Fig. 2.10 The participation of proline in a peptide bond.

H_2N–Tyr–Gly–Gly–Phe–Met–COOH
Enkephalin

Fig. 2.11 The structure of a typical peptide.

IDENTIFICATION OF PEPTIDE BONDS

The presence of a peptide bond is usually determined by the biuret reaction. Biuret has the formula $NH_2CONHCONH_2$ and is a simple substance possessing a peptide bond. When biuret is treated with $CuSO_4$ in alkaline solution, a purple colour is produced. This is known as the biuret reaction and, as expected, proteins give a strong reaction.

IONIC PROPERTIES OF PEPTIDES

THE NATURE OF THE CHARGED R-GROUPS

The ionizable, dissociable α-amino and α-carboxyl groups of the amino acids are blocked by peptide formation, except for the terminal residues. The ionized state of a protein therefore depends almost entirely on the R-groups; this, in effect, means those on aspartic and glutamic acids, lysine, arginine and histidine. This is illustrated in Fig. 2.12, which shows the structure of a hypothetical peptide containing all these groups. The numbers indicate the p*K* range of each dissociating group. As indicated above, histidine is very important because its charge can vary over the physiological pH range.

THE ISOELECTRIC POINTS OF PROTEINS

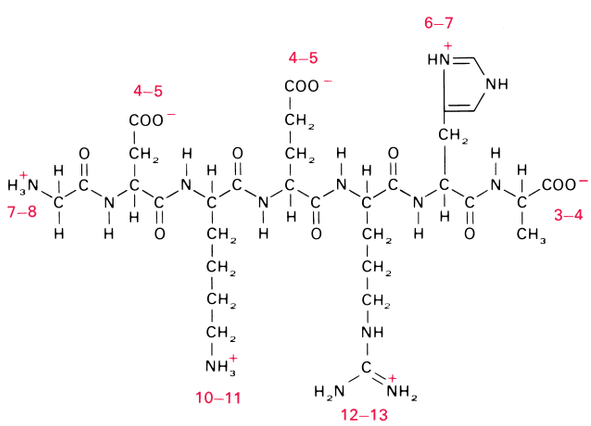
The isoionic point is the pH that results when the protein, freed of all other ions, is dissolved in water. The isoelectric point is the pH at which there is zero migration in an electric field (see below) to either electrode. The isoelectric points of a range of proteins are shown in Table 2.1. On the basis of these values, proteins are described as basic, neutral or acidic, depending on whether their overall charge at physiological pH is positive, approximately zero or negative.

ELECTROPHORESIS OF PROTEINS

Just as amino acids can be separated by electrophoresis so can proteins. Figure 2.13 shows the result of the electrophoresis of human serum proteins on a cellulose strip (paper can also be used) in a buffer at pH 8.6. The separated protein bands are visualized after staining with dye, and a densitometric scan provides an indication of the relative amount of protein in each

Table 2.1 The isoelectric points of some common proteins

Protein	Isoelectric point
Blood proteins	
α_1-Globulin	2.0
Haptoglobin	4.1
Serum albumin	4.7
γ_1-Globulin	5.8
Fibrinogen	5.8
Hemoglobin	7.2
γ_2-Globulin	7.4
Miscellaneous proteins	
Pepsin	1.0
Ovalbumin	4.6
Insulin	5.4
Histones	7.5–11.0
Ribonuclease	9.6
Cytochrome c	9.8
Lysozyme	11.1



glycyl-aspartyl-lysyl-glutamyl-arginyl-histidyl-alanine

Fig. 2.12 Polypeptides possess ionic properties, mainly due to the R-groups on the amino acid residues.

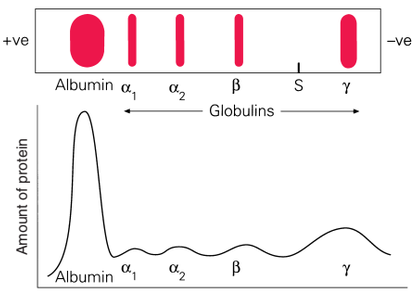


Fig. 2.13 Separation of serum proteins by electrophoresis on cellulose acetate.

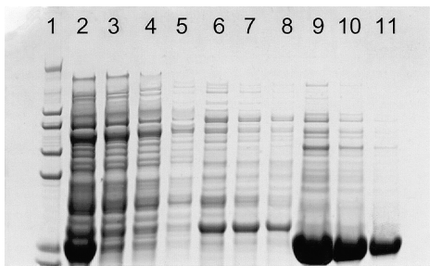


Fig. 2.14 An example of the use of SDS PAGE for following the purification of an enzyme from *Pseudomonas aeruginosa* expressed from a plasmid inserted into *E. coli* (see Recombinant DNA p. 49). Lane 1, marker proteins of various M_r down to about 30 K; lane 2, total proteins expressed from the plasmid in the presence of the inducer IPTG; lanes 3–11, the purification of the desired protein by the use of an affinity chromatography column: lane 3, eluent (washing) from the column; lane 4, low salt eluate, first fraction; lane 5, as for lane 4 but later fraction; lanes 6–11, proteins which were eluted from the column by imidazole; lanes 6, 7, 8, eluates from column with 20 mM imidazole; lanes 9, 10, 11, eluates from column with 200 mM imidazole; lane 11, the homogeneity of the desired protein (dimethylarginine dimethylamino hydrolase).

band. 'S' indicates the point of application of the serum before applying the current with the charges shown. Although the mobility of the proteins depends mainly on their relative charge, the size of the proteins also plays a part and this certainly contributes to the position of the large γ-globulin band. Although the serum proteins give the appearance of being separated into discrete bands, it should be remembered that, with the exception of serum albumin, each band contains many different proteins.

POLYACRYLAMIDE GEL ELECTROPHORESIS (PAGE) OF PROTEINS

Polyacrylamide gel can be used instead of cellulose acetate or paper for the separation of native proteins. Such a gel is commonly used in the presence of sodium dodecyl sulfate (SDS). In this case, oligomeric proteins (those composed of several discrete polypeptides) are separated in the form of their subunits.

Figure 2.14 shows the resolution of proteins by SDS-PAGE. The proteins are suspended in a 1% solution of SDS. This detergent disrupts most protein–protein and protein–lipid interactions. Very often, 2-mercaptoethanol is also added, to disrupt disulfide bonds. The electrophoretic mobility of most proteins, but not glycoproteins, depends on their size, as the negative charge contributed by SDS molecules bound to the protein is much larger than the net charge of the protein itself. A pattern of bands appears when the gel is stained with Coomassie Blue.

Agarose can be used in place of polyacrylamide gel for larger proteins or to obtain a different type of separation in the absence of SDS.

Two-dimensional PAGE can also be carried out, using different conditions in each direction: for example, an immobilized pH gradient (pH 4–7) in one direction and an 11–14% polyacrylamide gradient in the other.

NON-COVALENT BONDS IN PROTEINS

The distribution of charged amino acid groups in a polypeptide chain has already been described. The charged groups are important in terms of the folding of the chains because negatively charged groups will repel each other, as will positively charged groups, whereas closely positioned negative and positive charges will attract each other. There are, however, several other important interactions between the R-groups in proteins. These are illustrated in Fig. 2.15.

The ionic interactions already referred to are also known as salt bridges and are illustrated by an interaction between glutamate and arginine. The S–S bonds formed by the oxidation of two sulfhydryl groups are covalent and are particularly likely to be present in proteins when the physiological environment is unfriendly; they enhance the rigidity of the protein. An example is the proteins in the digestive secretions of the pancreas.

The other interactions are described as non-covalent and can be either apolar (i.e. hydrophobic) or polar (i.e. ionic and hydrogen bonding). Hydrophobic interactions result from: (i) van der Waals interactions, which arise from an attraction between atoms due to fluctuating electric dipoles originating from the electronic cloud and positive nucleus; (ii) the hydrophobic effect, which is the tendency of non-polar groups to associate with one another rather than to be in contact with water. Hydrogen bonds arise because, when a hydrogen atom is linked to an oxygen atom, there is a shift of electrons leading to a partial negative charge on the other atom. This produces an electric dipole that can interact with dipoles that exist elsewhere. The most common hydrogen bond is between –N–H and –C=O, as in the α helix and β-pleated sheet (to be described in Chapter 4), but other bonds are possible, as shown in Fig. 2.15.

PURIFICATION OF PROTEINS AND DETERMINATION OF RELATIVE MOLECULAR MASS

PURITY AND HOMOGENEITY

The purification of small molecules has traditionally ended with crystallization and the determination of various physical parameters, such as the melting point, but these procedures are much less applicable to macromolecules such as proteins. Even if proteins are crystallized, they might be contaminated by other proteins, by viruses or by other infective agents such as prions (see p. 60). The objective in the purification of proteins, therefore, is to produce a product that is homogeneous by all known criteria, which usually includes

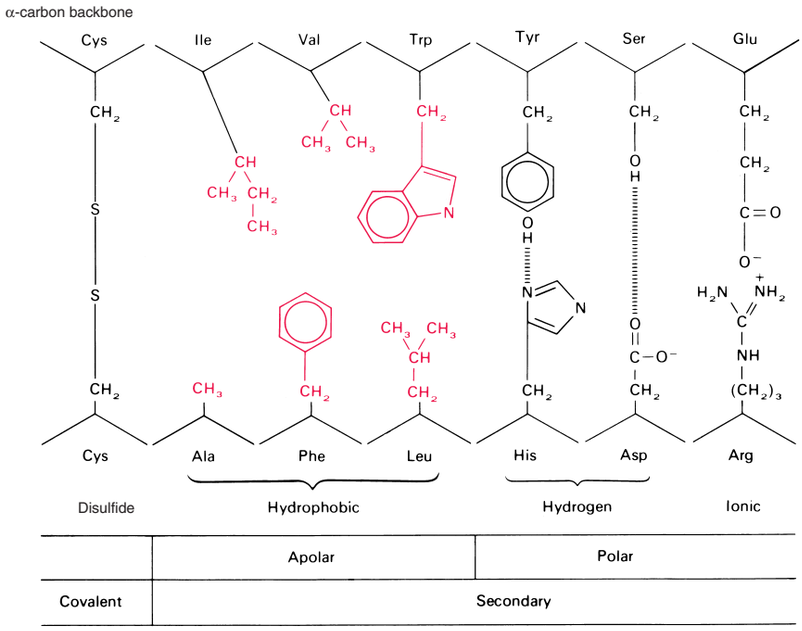


Fig. 2.15 The creation of non-covalent bonds is important in the formation of the tertiary structure of proteins. The various bonds are illustrated.

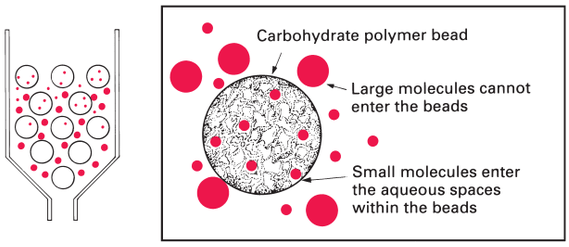


Fig. 2.16 Gel-permeation chromatography can be used for the removal of small molecules from protein solutions and for the separation of proteins according to their size.

electrophoresis under various conditions. To achieve this, many different methods are used, based on the characteristic properties of proteins, and in particular their ability to interact specifically with small molecules. The methods used must not impair the structure of the native protein or affect its biological activity. Traditional methods involved the differential solubility of proteins in solutions of ammonium sulfate, but many other methods are now available, such as gel-permeation chromatography, ion-exchange chromatography (similar to that already described for amino acids but using cellulose or Sephadex rather than a resin) and affinity chromatography. Some of these methods are described below.

GEL-PERMEATION CHROMATOGRAPHY

This technique utilizes a matrix based on dextran. This cross-linked polymer of dextran forms a mesh that can be penetrated only by molecules of a certain size (the greater the cross-linking, the smaller the holes of the mesh). The trade name of the dextran is Sephadex; various grades of Sephadex are produced and these differ in the extent of cross-linking. The principle of the method is shown in Fig. 2.16. Large molecules, which penetrate the mesh less readily, have less volume through which to permeate and thus elute more quickly. The matrix is normally packed in a column. The method can be used to separate small molecules, such as salts, from larger molecules, such as proteins, and to separate macromolecules of different sizes. Other materials such as agarose and Sepharose (a proprietary agarose) can be used as the basis of the matrix.

The ability of proteins to bind specifically to other molecules is the basis of affinity chromatography. In this technique ligand molecules that bind to the protein of interest are covalently attached to the beads in the form of a column. The ligands can be enzyme substrates or antibodies. The proteins are eluted by adding an excess of the ligand or by changing the salt concentration or pH of the elutent. (See p. 78 for immunoaffinity chromatography.)

NOMENCLATURE FOR THE SIZE AND DENSITY OF MACROMOLECULES

Formerly, the size of a molecule was described in terms of its molecular weight, but the term relative molecular mass (abbreviation M_r) is now preferred. Both M_r and molecular weight are ratios and hence it is incorrect to give them units such as daltons (symbol Da). It is thus incorrect to state that 'the M_r or the molecular weight of substance X is 10^5 Da'; the correct usage is 'M_r = 10 000'. The dalton is a unit of mass equal to one-twelfth the mass of an atom of carbon-12. Hence, it is correct to say that 'the molecular mass of X is 10^5 Da' or to use expressions such as 'the 16 000-Da peptide'. For entities that do not have a definable M_r, it is correct to state, for example, 'the mass of a ribosome is 10^7 Da'. A kilodalton (symbol kDa) is equal to 1000 Da.

Gel permeation can be used for the determination of the M_r of a protein. Plots of the elution volumes (V_e) of native proteins of known M_r on Sephadex G-75 and G-100 versus log M_r are shown in Fig. 2.17.

LARGE-SCALE SEPARATION OF PROTEINS

The scheme illustrated in Fig. 2.18 shows some of the many methods that are used for the separation of the plasma proteins. Cryoprecipitation depends on the lesser solubility of some proteins in the cold. DEAE (diethylaminoethyl)-, QAE (quaternary aminoethyl)- and SP (sulfopropyl)-Sephadex (or Sepharose) provide separation by ion exchange, as does CM (carboxymethyl)-Sepharose. Although every effort can be taken to produce a product that consists only of the protein of interest, purity cannot be guaranteed. Thus the isolated protein might be contaminated with very small amounts of other substances. Examples of such contamination are: (i) the virus that causes AIDS–HIV (see p. 29) in preparations of factor VIII, which is used in the treatment of people with hemophilia;
(ii) the presence of the factor that causes

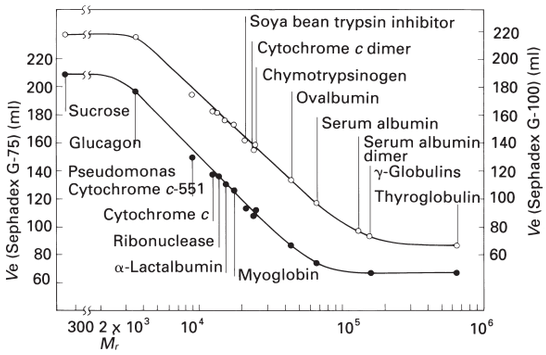


Fig. 2.17 Gel-permeation chromatography can be used to determine the M_r of a protein.

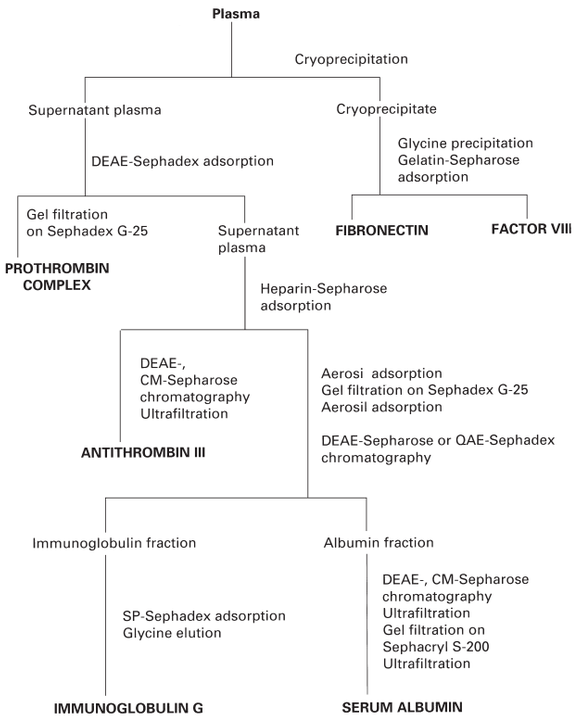


Fig. 2.18 Methods that can be used for the large-scale fractionation of proteins.

Creutzfeldt–Jakob disease (see p. 60) in preparations of human growth hormone; (iii) and the virus that causes hepatitis C in products from blood. Some of these contaminants can be inactivated by heat treatment. In many cases, the alternative of expressing a recombinant DNA for the chosen human protein in a vector such as *E. coli* or yeast (see p. 49) is to be preferred, but care must be taken to eliminate the proteins of the vector from the human protein preparation. Such methods are used for the preparation of erythropoietin, which is used for the treatment of anemia in patients with kidney failure. Serum albumin cannot as yet be obtained in this way.

THE DETERMINATION OF THE AMINO ACID SEQUENCE OF PROTEINS

Proteins have precisely defined amino acid sequences and there are many reasons for wishing to know this sequence for each protein. As shown later (p. 34), it might be possible to achieve this by an indirect method after determining the structure of the gene for the protein and deducing the amino acid sequence from knowledge of the genetic code. Direct methods involve the determination of the N-terminal amino acid followed by Edman degradation. Because this method is limited to about 50 amino acids, it is first necessary to break larger proteins into smaller polypeptides, either chemically or by the use of proteolytic enzymes (see p. 154). Provided the peptides overlap, it is possible to deduce the sequence of the entire protein. Edman degradation involves the reaction of phenylisothiocyanate with the N-terminal amino acid and its release by mild acid. The procedure is continued in a stepwise, automated manner.

PROTEIN STRUCTURAL HIERARCHIES

The polypeptide chains of proteins fold in various ways, both within chains and with other chains. This folding is essential for the biological activity of proteins and it is this intricate folding that must be preserved during the procedures involved in protein purification. Although it has long been claimed that the manner of folding of the polypeptide chains is determined solely by the amino acid sequence of the chains, it is now accepted that proteins with identical amino acid sequences can exist in differently folded forms, and that such folding can be influenced by the presence of other proteins, known as molecular chaperones (see p. 15).

It is useful to consider protein structure in terms of the four hierarchies shown in Fig. 2.19.

PROTEIN DENATURATION AND RENATURATION

A protein that possesses its own unique biological property is known as a native protein, to distinguish it from a protein that has lost this property and which is described as denatured. A denatured protein has lost its three-dimensional structure, also known as its conformation. Denaturation can be either irreversible or reversible. An example of irreversible denaturation is the application of heat when an egg is boiled; the egg white (albumen) coagulates in an irreversible manner. In fact, this is a common event during cooking that renders proteins more susceptible to the action of proteolytic enzymes when the food is eaten.

Reversible denaturation can be achieved by the careful use of reagents such as urea and mercaptoethanol. Urea destroys the water structure and hence decreases the hydrophobic bonding of the R-groups of the amino acid residues (see Fig. 2.15), resulting in the unfolding and dissociation of the protein molecules. Mercaptoethanol reduces the S–S bonds.

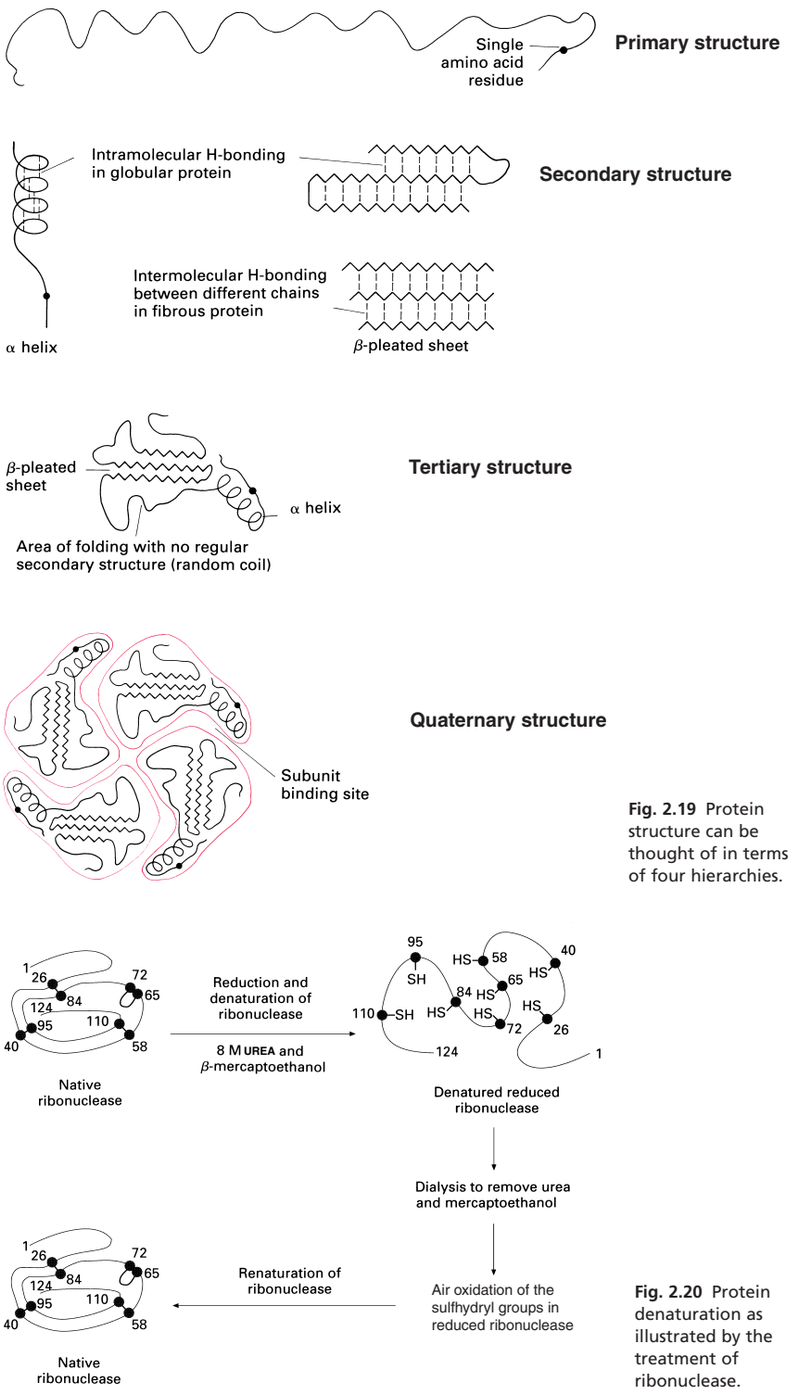


Fig. 2.19 Protein structure can be thought of in terms of four hierarchies.

Fig. 2.20 Protein denaturation as illustrated by the treatment of ribonuclease.

It might therefore be possible to renature the protein when the urea and mercaptoethanol are removed. These processes are shown in Fig. 2.20 for ribonuclease.

Renaturation has been taken to indicate that a protein with the 'correct' primary structure will fold spontaneously to give the unique structure required for biological activity. This process is termed 'protein self-assembly'. It is now realized that there are two means whereby renaturation can be assisted. One involves the enzyme protein disulfide isomerase, an enzyme that plays a role 'correcting' wrongly paired S–S bonds. The other involves molecular chaperones, which have already been referred to. These can be defined as a family of unrelated classes of proteins that mediate the correct assembly of other polypeptides but are not components of the functional assembled structures. Examples are heat-shock proteins synthesized by cells after their exposure to an abnormally increased temperature.

PEPTIDES, STRUCTURE AND BIOLOGICAL ACTIVITY

EXAMPLES OF SMALL PEPTIDES

There are many naturally-occurring peptides with a wide range of activity, such as hormones, first messengers in neurotransmission, local mediators and antibiotics. These peptides vary in length from the three amino acids of thyrotropin-releasing hormone (TRH) to the 231 amino acids of human gonadotropin. Even the smallest peptides have a very specific activity.

The structures of some typical peptides are shown in Fig. 2.21. The N- and C-termini are often modified. Thus, in TRH, the N-terminus is a cyclized glutamic acid (pyroglutamic acid) and there is an amide at the C-terminus. It is possible that such modifications enhance metabolic stability by protecting the peptides against exopeptidases.

Examples of small peptide hormones produced in the posterior pituitary are oxytocin and vasopressin. The structures of these are shown in Fig. 2.21; again, the C-terminus is an amide. The vasopressins are more correctly named antidiuretic hormone (ADH), because their most important physiological action is to promote reabsorption of water from the distal renal tubule. Oxytocin accelerates birth by stimulating contraction of uterine smooth muscle. These structures illustrate

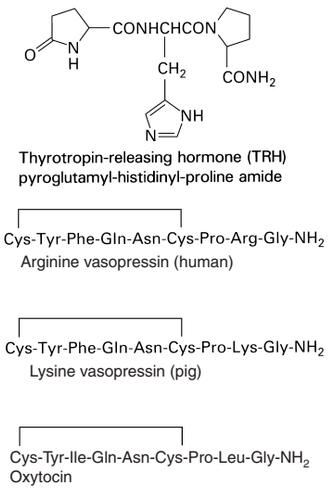


Fig. 2.21 Structures of some typical peptides.

the specificity of peptides in that small changes of structure are associated with major functional change. Many of the antibiotic peptides are cyclized.

EXAMPLES OF LARGER HORMONES

Somatotropin (growth hormone) and prolactin are protein hormones of the anterior pituitary; lactogen is produced by the placenta. All three hormones are closely related in structure.

Another group of hormones is a family of glycoproteins, which includes thyrotropin, follicle-stimulating hormone (FSH) and chorionic gonadotropin. These compounds all contain numerous N-linked branched carbohydrate chains – hence the name of the group.

Some peptide hormones are first synthesized as larger peptides, which are subsequently split in the tissues into smaller peptides with discrete activities. Such large precursor peptides are called polyproteins. Good examples are the adrenocorticotropin (ACTH) peptides produced from proopiocorticotropin (see p. 33).

THE USE OF THE MASS SPECTROMETER IN PROTEIN STRUCTURE STUDIES

There have been important developments in the approach to protein structure determination in recent years. In particular, techniques in mass spectrometry have been introduced to rapidly identify proteins and to enable the determination of polypeptide amino acid sequence. For identification, the protein is initially digested with trypsin (or similar enzyme) and the peptides formed simultaneously analysed with very high sensitivity by matrix-assisted laser desorption mass spectrometry. The list of peptide masses are then compared *in silico* (i.e. by computer) against the computed masses of all known proteins in the international databases. This technique of peptide fingerprinting usually results in rapid identification of the original protein.

There are also approaches to obtain amino acid sequence information directly from proteins but, more generally, sequence is derived on smaller polypeptides and peptides from a protein by collision-induced dissociation mass spectrometry. In this case the molecular ions of the peptides are collisionally fragmented at their constituent peptide bonds and the sequence deduced from the resulting mass spectrum – a process that also can be carried out automatically by computer. The combination of peptide fingerprinting and sequencing has led to a rapid and sustained expansion in studies of the proteome, in which complex mixtures of proteins are initially separated by two dimensional electrophoresis followed by in-gel digestion and mass spectrometric analysis.